

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



A study of parapinopsin in amphibians

Zainab, Tayyaba

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

A STUDY OF PARAPINOPSIN IN AMPHIBIANS

Tayyaba Zainab

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy at King's College London

The Institute of Pharmaceutical Science,
School of Biomedical and Health Sciences,
King's College London,
Hodgkin's Building,
Guy's Campus,
London
SE1 1UL

ABSTRACT

Parapinopsin (PNP) is a novel pigment protein present in pineal gland of the brain. PNP protein is similar in function to the human visual pigment protein rhodopsin, found in the eye. The aim of project is to sequence and characterize PNP in different frogs species to understand more about PNP as little is known about its structure and function.

Tadpole cDNA from different frogs species (*X.laevis*, *Rana temporaria* and *Rana tigrina*) were used during the project. PCR primers were designed against *Xenopus tropicalis* PNP cDNA the only reported sequence at NCBI. PNP PCR products were cloned into pGEM vectors and sequenced. The cDNA sequencing results of *X. laevis* showed 100% homology but *R. temporaria* and *R. tigrina* showed ~99.9% similarity with *X. tropicalis*. The 0.1% dissimilarity is due to a point mutation from G to A at the 2nd position in the 156th codon. Genomic DNA sequence study results revealed a heterozygosity in *R. temporaria* and a point mutation in *R. tigrina* PNP. This point mutation is very important as it creates a stop codon possibly resulting in protein truncation. These findings were further confirmed by SSP-PCR (sequence specific primers- PCR) results. Western blotting was used to determine more about the size of the protein produced. Two peptides regions were selected at N and C termini. Western Blotting results support the cDNA study results with a 37 kDa protein band of PNP in the *X. laevis* protein sample only. Bioinformatics softwares were used to study evolutionary genetics and to predict the 2D structural elements and a 3D structural model for PNP from *X. laevis*. Further studies were conducted on *X. laevis* melanophores to look for PNP expression. Northern Blotting results indicated the absence of PNP in melanophores. For protein expression studies the entire coding sequence of PNP gene was chemically synthesized and subsequently cloned into the vector pASKGPCR for expression in *E. coli*. The aim of this study was to overexpress and purify the protein for functional studies as well as structural studies using X-ray crystallography and high field NMR.

ACKNOWLEDGEMENT

Millionth gratitude to Allah Almighty, The Most Beneficent and The Most Merciful who bestowed me with the potential to seek knowledge and to explore some of the many aspects of His creation. Countless blessings and clemencies of Allah may be upon our Holy Prophet (S.A.W.W) and His Ahl-e-Bait (A.S.), the fortune of knowledge, who took the humanity out of the abyss of ignorance and elevated it to the zenith of consciousness.

I feel privileged and have great pleasure to place on record my sincere indebtedness to my learned supervisor Dr. Peter Eagles under whose sincere guidance and obliging supervision this work was completed. His academic worth is enormous and I have tried my level best to gain knowledge from his as much as possible. Yet I feel no shame to admit that best put efforts of mine were still insufficient in comparison to endless realms of his knowledge. I shall be failing in my duty if I do not put across my thanks to my co-supervisor Dr. Mark Sanderson for providing materials and valuable suggestion whenever I needed them.

I would like to express my heartiest gratitude and thanks to Dr. Amer Quershi for helping me to start my work at KCL and helping me to settle down in the lab. Countless thanks for his magnanimous cooperation and valuable suggestions during conduct of this research. My special thanks go to Dr. Atsuko, Dr. Fiona Howell and Praven for their indispensable support during Western blotting studies.

Last but not the least I would like to thank Higher Education Commission (HEC) for funding the project "Establishment of University Institute of Biochemistry and Biotechnology", to Pir Mehr Ali Shah University of Arid Agriculture Rawalpindi, Pakistan. Also International Hardship Fund, King's College London, Bestway Foundation, UK and Charles Wallace Trust, UK for providing me the funding when needed it the most.

I dedicated this thesis to my Parents and in-laws, my loving Daughter, caring husband and the whole family.

CONTENTS

ABSTRACT	1
ACKNOWLEDGEMENT	3
CONTENTS	5
Index of Tables	12
Index of Figures	14
List of Abbreviations	12
Chapter 1 Introduction	22
1.1 Structure of the Eye	22
1.2 Structure of Retina	23
1.3 Function of the eye	24
1.4 Brain	24
1.4.1 The Frog Brain	25
1.5 The Vertebrate Pineal organ	26
1.5.1 Pineal and parapineal organs of Lamprey and Teleost	26
1.5.2 The Pineal Complex of Amphibians	27
1.5.3 The Pineal Organ in Reptiles	30
1.5.4 The Mammalian Pineal gland	31
1.6 Circadian Rhythm	31
1.6.1 Circadian Rhythm in Vertebrates	33
1.6.2 Molecular Circadian Clock	34
1.7 Photosensory Organs	36
1.7.1 Visual Photoreceptors	36
1.7.2 Non-visual Photoreceptors	36
1.8 Classification of Photoreceptors	36
1.8.1 Ciliary Photoreceptors Cells	37
1.8.2 Rhabdomeric Photoreceptors Cells	37

1.8.3 Vertebrates Ciliary Opsins	37
1.9 GPCR: The Opsins	38
1.9.1 The Vertebrate Opsin/encephalopsin Subfamily	39
1.9.1.1 Vertebrate opsins	39
1.9.1.2. The Vertebrate Non-Visual Opsin	40
1.9.2 Encephalopsins	42
1.9.4 Gs Opsins	42
1.9.5 Gq Opsins	42
1.9.4 Photoisomerases	43
1.10 Molecular Machinery Involves in Light Sensitive Visual System	43
1.10.1 Light Switch Molecules	43
1.10.2 Rhodopsin.....	45
1.10.2.1 The Chromophore	47
1.10.3 Classification of Animal Phototransduction Cascade	47
1.10.3.1 Phosphoinositol Signalling (rhabdomeric type photoreceptor cells)	48
1.10.3.2 Cyclic Nucleotide Signalling (Cyclic type photoreceptor cells)	50
1.10.3.3 Types of Photopigments.....	54
1.10.3.3 The counterion.....	55
1.11 Parapinopsin.....	55
1.11.1 Comparison of Distinguishing Properties of Parapinopsin and other Opsins	56
1.11.2 Evolution in Position of Counterions and Parapinopsin.....	58
1.11.3 Visual-Arrestin and β -Arrestin	59
1.11.4 Monostable Vs Bistable Pigments G Protein Activation	59
1.12 Work Done on Parapinopsin.....	61
1.12.1 Catfish Parapinopsin	61
1.12.2 Lamprey Parapinopsin.....	61
1.12.2.1 Types of Ganglion cells in Pineal organ	61
1.12.3 β Arrestin and Lamprey Pineal	61
1.12.4 Pineal the UV Reception Organ in Vertebrates.....	61

1.12.5 Expression of Parapinopsin in Reptile Lineage	62
1.12.6 Position of Parapinopsin in Phylogenetic Tree.....	62
1.13 Amphibian Species Used during Current Studies & their Genetics	64
1.13.1 Classification of Frogs	65
1.14 Aim of Project	66
CHAPTER 2 MATERIAL & METHODS.....	68
2.1 RNA Extraction and DNA Purification	68
2.2 Quantification of DNA and RNA Concentration	68
2.3 Synthesis of cDNA by RT	69
2.4 Primer Designing	69
2.5 Polymerase Chain Reaction	71
2.6 Sequence Specific Primer- Polymerase Chain Reaction (SSP-PCR).....	72
2.6.1 SSP-PCR Primer Sequence.....	73
2.7 Visualization of PCR products	74
2.8 Precipitation of PCR Products	75
2.9 DNA Ligation.....	75
2.10 Preparation <i>E. coli</i> Competent Cells	76
2.11 Transformation	76
2.12 Plasmid DNA Purification, Mini preparation	77
2.13 Glycerol Stocks.....	77
2.14 DNA Sequencing	77
2.15 AT Cloning	77
2.16 Tadpole Anaesthesia	79
2.16.1 Preparation of Brain Tissue Lysates and Protein Assay	79
2.17 Western Blot Sample Preparation	81
2.18 SDS-polyacrylamide Gel Electrophoresis	82
2.19 Western Blot Analysis.....	83
2.20 Petri Dish Culture of Frog's/ toads Embryos	85
2.21 Culturing of free Swimming Tadpoles	85

2.22 Tadpoles Preservation.....	86
2.23 Cell Culture Work	86
2.23 RNA Sample Preparation and RNA Denatured Gels	87
2.24 Northern Blotting.....	90
2.25 Protein Expression studies	90
2.25.1 Parapinopsin Gene Synthesis	90
2.25.2 Expression Vectors.....	90
2.25.3 Primers used during Expression Studies	91
2.25.4 Cloning of the Parapinopsin gene in pASKGPCR.....	92
2.25.6 Co- Expression of pASKGPCR and FtsH.....	93
Chapter 3 Sequencing of Parapinopsin in Frogs	95
3.1 Aim of Project	95
3.2 Sequencing of <i>Xenopus tropicalis</i> Parapinopsin cDNA	95
3.2.1 PCR amplification of Parapinopsin cDNA of	95
3.3 Sequencing of <i>Xenopus Laevis</i> Parapinopsin cDNA	97
3.3.1 PCR Amplification of <i>Xenopus Laevis</i> Parapinopsin cDNA	97
3.3.2 AT Cloning of the PNP PCR Products in the pGEM® T Easy Vector System.....	99
3.4.3 Sequence Analysis of the PNP1 products.....	103
3.5 PCR amplification of <i>Rana Temporaria</i> Parapinopsin cDNA	106
3.5.1 AT Cloning of the R. temporaria PNP PCR Products in the pGEM® T Easy Vector System	108
3.5.2 Sequence Analysis of the PNP1 &PNPd products for <i>R. temporaria</i>	110
3.5.3 Genotype and Quality value calculation for R. temporaria cDNA peak at 2nd position of the 156th codon	115
3.6 <i>Rana Tigrina</i> Parapinopsin cDNA study.....	116
3.6.1 Sequencing of <i>Rana tigrina</i> Parapinopsin cDNA.....	117
3.7 Parapinopsin Genomic DNA study.....	118
3.7.1 Exon 2 Sequencing of <i>Xenopus Laevis</i> Parapinopsin gDNA.....	120
3.7.2 Sequencing of <i>Rana temporaria</i> Parapinopsin gDNA.....	121
3.7.3 Genotype and QV calculation for <i>R. temporaria</i> gDNA at 2nd position of 156th codon	126

3.8 Sequencing of artificial heterozygotes (cloned <i>X. laevis</i> and <i>R. temporaria</i> cDNA mixture).....	127
3.8.1 Genotype and QV calculation for “artificial heterozygotes” at the 2nd position of 156th codon	129
3.9 <i>R. temporaria</i> gDNA Single Tadpole Study	130
3.9.1 Genotype and QV Calculation for <i>R. temporaria</i> gDNA at 2nd Position of 156th Codon	133
3.10 Sequencing of <i>Rana tigrina</i> Parapinopsin gDNA	134
3.10.1 Genotype and QV calculation for the 2nd position of 156th codon in Parapinopsin from <i>R. tigrina</i> gDNA	137
3.11 Summary of cDNA and gDNA Parapinopsin study Results	139
3.12 Summary of Genotype at 2 nd position of 156 th codon.....	140
3.13 Discussion	142
Chapter 4 Detection of Single Nucleotide Polymorphism by the use of Sequence Specific Primers using PCR (SSP-PCR)	149
4.1 Aim of Study	149
4.2 SSP- PCR Primer Designing	149
4.3 SSP-PCR amplification of <i>X.laevis</i> Parapinopsin gDNA	151
4.4 SSP-PCR Amplification of <i>Rana Temporaria</i> Parapinopsin gDNA	152
4.5 SSP-PCR Amplification of <i>R.tigrina</i> Parapinopsin gDNA.....	153
4.6 SSP-PCR Amplification of Artificial Heterozygotes	154
4.7 Summary of SSP-PCR Results	155
4.8 SSP-PCR Vs Sequencing Results	156
Chapter 5 Study of Parapinopsin Protein	159
5.1 Aim of Study.....	159
5.2 Study of Parapinopsin Protein: Peptide Design	159
5.2.1 Parapinopsin Sequence Study and Selection of Epitopes for the Generation of Polyclonal Antibodies	161
5.2.2 Secondary Structure Prediction of Parapinopsin Protein	162
5.2.3 Production of Polyclonal Antibodies and Western Blot analysis	163
5.2.3.1 Antigenic Epitope Selection.....	163
5.2.4 Parapinopsin Protein 3D Structure Prediction.....	165

5.2.5 <i>R. temporaria</i> Parapinopsin Protein 3D Structure	166
5.3 Gel staining and Parapinopsin protein analysis	167
5.4 Western Blot analysis for <i>X. laevis</i> & <i>R. temporaria</i> Parapinopsin Protein using N and C Terminal Antibodies	169
5.4.1 Western Blotting analysis for <i>X. laevis</i> & <i>R. tigrina</i> Parapinopsin Protein using N and C terminal Antibodies	172
5.5 Discussion	174
6 Parapinopsin Study in Melanophores by Northern Blotting	178
6.1 Aim of Study	178
6.2 Northern Blotting Steps	178
6.2.1 <i>X. laevis</i> Melanophore mRNA Extraction & Preparation of Denatured Agarose gel	178
6.2.2 Cross-Linking of mRNA to a Nylon Membrane	179
6.2.3 DIG Labelled Single Stranded RNA Probe Synthesis	179
6.3 Northern Blot Results	180
6.4 Summary of Northern Blotting Results	180
6.5 Discussion	181
Chapter 7 Expression of Parapinopsin Protein	183
7.1 Aim of Study	183
7.2 Tailoring of the Chemically Synthesized Parapinopsin Gene	183
7.2.1 Two Step PCR	183
7.2.2 Restriction Digestion of Parapinopsin gene	184
7.2.3 pASK75-GPCR Plasmids	184
7.2.4 Cloning of the Parapinopsin in the pASK75 plasmids	185
7.2.5 Sequence Analysis of the pASK-GPCR-GFP	188
Chapter 8 Bioinformatics Based Study of Parapinopsin	189
8.1 Aim of Study	189
8.2 Protein Sequences used for Multiple Sequence Alignment	189
8.2.1 CLUSTAL W 2.1 Multiple Sequence Alignment	189
8.2.2 Multiple Sequence Alignment Result	192

8.3 Parapinopsin Protein Phylogenetic Tree	196
8.3.1 Multiple sequence comparison by log-expectation (MUSCLE)	197
8.4 Parapinopsin Study by Basic Local Alignment Similarity Tool (BLAST)	210
8.5 BLAST Results Phylogenetic Analysis by Neighbour Joining Method	213
8.6 Discussion	214
Chapter 9 Discussion	217
9.1 Discussion	217
9.2 Further Work.....	217
9.3 Conclusion	221
9.4 Future Development	222
References	223
Appendix.....	22346

Index of Tables

Table 1.1 classification and functions of vertebrate opsins	40
Table 1.2 Vertebrates non-visual opsins	42
Table 1.3 Types of photopigments	54
Table 1.4 Types and functions of arrestin	59
Table 1.5 Information about Frog species used during current studies	64
Table 1.6 Classification of <i>X. tropicalis</i> , <i>X. laevis</i> , <i>R. temporaria</i> and <i>R. tigrina</i>	65
Table 2.1 Formulae used to measure RNA and DNA concentrations.....	68
Table 2.2 Primer sets used within the project, 5'→3'.....	71
Table 2.3 (a) Recipe for PCR Master Mix (25µl reaction mixture)	71
Table 2.3 (b) Thermo cycler's PCR program for 30 cycles	72
Table 2.4 SSP-PCR Primer sets used within the project, 5'→3'.....	74
Table 2.5 Recipes of lysis buffer and the sample loading buffer.....	80
Table 2.6 N and C terminus Peptide selection criteria	81
Table 2.7 Recipes of the Buffers used during SDS-PAGE electrophoresis and Western.....	84
Table 2.8 Modified Barth's saline (MBS)	85
Table 2.9 MR (Modified Frog Ringer's) solution composition	86
Table 2.10 Buffer Recipes for RNA gels	89
Table 2.11 Primer sets used in expression studies, 5'→3'	91
Table 3.1 Sequencing results of the region around 156 th codon from various cDNA samples of <i>R. temporaria</i> Parapinopsin	114
Table 3.2 Genotype and QV % for <i>R. temporaria</i> cDNA at 2nd position of 156th codon	115
Table 3.3 Sequencing results around 156 th codon of <i>R. tigrina</i> cDNA Parapinopsin.....	118
Table 3.4 Sequencing results around 156 th codon of gDNA of <i>X. laevis</i> Parapinopsin ..	121
Table 3.5 Polymorphic sequencing results around 156 th codon of Parapinopsin from 8 samples of <i>R. temporaria</i> gDNA.....	125
Table 3.6 Genotype and QV % for <i>R. temporaria</i> gDNA at 2nd position of 156th codon	126
Table 3.7 Sequencing results around the 156 th codon of 50:50 mixture of cloned cDNA from <i>X. laevis</i> and <i>R. temporaria</i>	128

Table 3.8 Genotype and QV % for “artificial heterozygotes” at 2nd position of 156th codon	129
Table3.9 Polymorphic sequencing results around 156 th codon of Parapinopsin gDNA from 5 samples of single tadpole <i>R. temporaria</i> gDNA	132
Table 3.10 Genotype and QV % for <i>R. temporaria</i> single tadpole gDNA at 2nd position of 156th codon	133
Table3.9 Sequencing results of gDNA study around 156 th codon of <i>R. tigrina</i> Parapinopsin	136
Table 3.10 Genotype and QV % for at 2nd position of 156th codon of Parapinopsin from <i>R. tigrina</i> using gDNA	137
Table 3.11 Summary of Parapinopsin study in frogs.....	139
Table3.12: Genotype and Quality value calculation of peaks at 2 nd position of 156 th codon	141
Table4.1 <i>X.tropicalis</i> Parapinopsin cDNA sequence and SSP-PCR primer sequence..	150
Table4.2 Summary of SSP-PCR primer results according to genotype of tadpoles	155
Table4.3 Result summary of sequencing Vs SSP-PCR.....	156
Table 5.1 <i>X. tropicalis</i> Parapinopsin cDNA sequence and ORF information	161
Table 6.1 Summary of Northern Blot Results.....	181
Table 8.1 Multiple Sequence Alignment CLUSTAL W 2.1	191
Table 8.2 Multiple Sequence Alignment Result (MSA) Summary.....	195

Index of Figures

Figure1.1: Structure of the eye	23
Figure1.2: Structure of the Retina	24
Figure1.3 Frog's brain dorsal and ventral view	26
Figure 1.4 Diagrammatic representation of the Lamprey pineal organ	27
Figure 1.5 Diagrammatic presentation of the Pineal complex of the Frog	29
Figure1.6 Diagrammatic presentation of the Frog's pineal, parapineal organs and the retina	30
Figure1.7 (a) Cyclic overview of circadian rhythm.....	32
Figure 1.7(b) Schematic representations of steps involved in melatonin production	33
Figure1.8 Steps involved in vertebrate melatonin synthesis	34
Figure1.9 Feedback mechanism of circadian clock	35
Figure 1.10 Diagrams of rhabdomeric and ciliary photoreceptor cells	38
Figure 1.11 Schematic representation of the opsins phylogenetic relationship	44
Figure 1.12 Bovine Rhodopsin Structure	46
Figure 1.13 The animal's opsin phototransduction cascade	48
Figure 1.14 Steps involved in phosphoinositol signalling	49
Figure 1.15 a The light and dark phototransduction cycle in the rod cell	51
Figure 1.15 b The cyclic activation of GPCR G-protein.....	51
Figure 1.16 Signalling pathways functional diversity	53
Figure 1.17 Binding of chromophore with Lys296 and position of counterions	55
Figure 1.18 Amino acid sequence comparison of rhodopsin and Parapinopsin	56
Figure 1.19 Important amino acid positions in Parapinopsin compared with other opsins	57
Figure 1.20 Conformational changes in helix VI and V in bovine rhodopsin and Parapinopsin	60
Figure 1.21 Position of Parapinopsin among other opsins in a phylogenetic tree	63
Figure 2.1 SSP-PCR technique	73
Figure 2.2 Map of the pGEM®T Easy Vector system.	79
Figure 2.3 Schematic presentation of the Northern Blotting mechanism	87

Figure 2.4 (a) Expression vector pASKGPCR.....	92
Figure2.4 (b) The 5' and 3' modification of the Parapinopsin gene	93
Figure2.5 Expression vector pASKGPCR- GFP (A. JAMES LINK <i>et al.</i> , 2008)	93
Figure2.6 pBAD33 vector	94
Figure 3.1 Electrophoretic Analysis of PCR reaction using PNPd on <i>X. tropicalis</i> cDNA	96
Figure 3.2 Electrophoretic Analysis of PCR reaction using β -actin primers on <i>X. laevis</i> cDNA.	97
Figure 3.3 Electrophoretic Analysis of PCR reaction using PNP1 (a) and PNPd (b) primer sets on <i>X. laevis</i> cDNA.	98
Figure 3.4 Electrophoretic Analysis of ligation reaction PCR	100
Figure 3.5 Electrophoretic Analysis of the PCR products of the pGEM vector Colony Screen	102
Figure 3.6 Electrophoretic Analysis of the PCR products of cloned pGEM PNP1 pGEM PNPd	103
Figure 3.7 <i>X. laevis</i> sequence matches with <i>X. tropicalis</i> Parapinopsin cDNA sequence	105
Figure 3.8 Electrophoretic Analysis of PCR reaction using β -actin primers on <i>R. temporaria</i> cDNA	106
Figure 3.9 Electrophoretic analysis of PCR reaction using PNP1 (a) and PNPd (b) primer sets on <i>R. temporaria</i> cDNA.....	107
Figure 3.10 Electrophoretic Analysis of ligation reaction PCR	108
Figure 3.11 Electrophoretic Analysis of the PCR products of the pGEM vector Colony Screen	109
Figure 3.12 Electrophoretic Analysis of the <i>R. temporaria</i> cDNA PCR products of the cloned insert	110
Figure 3.13 <i>R. temporaria</i> sequence matches with <i>X. tropicalis</i> Parapinopsin cDNA sequence	112
Figure3.14 The percentage of QV for sequencing <i>R. temporaria</i> cDNA at the peak of the 2nd position of 156th codon	116
Figure 3.15 Electrophoretic Analysis of the <i>R. tigrina</i> cDNA PNP (O) PCR products....	117
Figure 3.16 Highlighted Exons in Zebrafish and <i>X. tropicalis</i> Parapinopsin cDNA Sequence.....	119
Figure 3.17 Electrophoretic Analysis of the <i>X. laevis</i> gDNA PNPexon2 PCR products.	120
Figure 3.18 Electrophoretic Analysis of the <i>R. temporaria</i> gDNA PNPexon2(I) (a) and PNP exon2(O) PCR products (b)	122

Figure3.19 The percentage QV for sequencing <i>R. temporaria</i> gDNA around the peak at the 2 nd position of the 156 th codon.....	127
Figure3.20 The percentage QV for sequencing of “artificial heterozygotes” at 2 nd .position of 156 th codon	130
Figure 3.21 Electrophoretic analysis of the single tadpole <i>R. temporaria</i> gDNA PNPexon2(I) (a) and PNP exon2(O) PCR products	131
Figure3.22 The percentage QV for sequencing of single tadpole <i>R. temporaria</i> gDNA at the 2 nd position of the 156 th codon for Parapinopsin	134
Figure 3.23 Electrophoretic Analysis of <i>R. tigrina</i> gDNA PNPexon2(I) & PNPexon2(O) PCR products	135
Figure3.24 The percentage QV for sequencing of <i>R. tigrina</i> gDNA at 2 nd position of 156 th codon for Parapinopsin.....	138
Figure 4.1 Electrophoretic analysis of PCR reaction using SSP1pnpA, SSP1pnpG forward and pnp reverse primers on <i>X.laevis</i> gDNA.	151
Figure 4.2 Electrophoretic Analysis of PCR products from <i>R.temporaria</i> gDNA with SSP1pnpA, SSP1pnpG, SSP2pnpTA, SSP2pnpTG forward primers and common pnp reverse primer.....	152
Figure 4.3 Electrophoretic analysis of PCR reaction using SSP1pnpA, SSP1pnpG forward and pnp reverse primers on <i>R.tigrina</i> gDNA.	153
Figure 4.4 Electrophoretic analysis of the PCR reaction using SSP1pnpA, SSP1pnpG forward and pnp reverse primers on artificial heterozygotes of cDNA.	154
Figure 5.1 <i>X. tropicalis</i> Parapinopsin protein sequence and selection of peptide regions	162
Figure 5.2 Predicted secondary structure of <i>X. tropicalis/ X. laevis</i> Parapinopsin protein	164
Figure 5.3 Predicted 3D structure of <i>X. tropicalis</i> Parapinopsin protein with selected N and C terminal peptides.....	165
Figure 5.4 The predicted <i>R. temporaria</i> Parapinopsin protein	166
Figure 5.5 (a) 15% SDS gel electrophoretic analysis of <i>X. laevis</i> , <i>R. temporaria</i> Parapinopsin, actin and haemoglobin protein samples.....	167
Figure 5.5 (b) 15% SDS gel electrophoretic analysis of <i>R. tigrina</i> Parapinopsin, actin and haemoglobin protein samples.....	168
Figure 5.6 Western Blot analysis for <i>R. temporaria</i> & <i>X. laevis</i> Parapinopsin protein with C and N- terminal PNP antibodies.....	172
Fig 5.7 Western Blot analysis for <i>R. temporaria</i> & <i>X. laevis</i> Parapinopsin protein with C and N- terminal Parapinopsin antibodies	173
Figure 6.1 Electrophoretic Analysis of <i>X.laevis</i> melanophores, <i>X.laevis</i> tadpoles and Kcl22 mRNA	178

Figure 6.2 Electrophoretic Analysis of the PCR product of cloned pGEM PNP1.....	179
Figure 6.3 Northern Blot analyses for Parapinopsin in <i>X. laevis</i> tadpole, <i>X. laevis</i> Melanophore and Kcl22 mRNA with <i>X. laevis</i> Parapinopsin cDNA DIG labelled probe.	180
Figure 7.1 Electrophoretic Analysis of the PCR reaction using ParaP2 forward & common ParaP reverse primer on the chemical synthesized <i>X. tropicalis</i> Parapinopsin cDNA...	184
Figure 7.2 Electrophoretic Analysis of restriction digested pASK-GPCR and pASK-GPCR-GFP.....	184
Figure 7.3 Electrophoretic Analysis of ligation reaction PCR.....	185
Figure 7.4 Electrophoretic Analysis of the PCR products of the pASK75-GPCR and pASK75-GPCR–GFP vectors Colony Screen	186
Figure 7.5 Electrophoretic Analysis of the PCR products of the pASK75-GPCR and pASK75-GPCR–GFP vectors Colony Screen	187
Figure 8.2 Parapinopsin phylogenetic tree with other opsins.....	196
Figure 8.4 Multiple Sequence Alignment (MUSCLE)	208
Figure 8.5 Parapinopsin DNA phylogenetic tree	209
Figure 8.6 (a) Graphical representation of <i>X. tropicalis</i> Parapinopsin BLAST results ..	210
Figure 8.6 (b) <i>X.tropicalis</i> BLAST results	212
Figure 8.7 Opsin based phylogenetic tree constructed by neighbour joining method ...	213

List of Abbreviations

2D	2 dimensional
3D	3 dimensional
7TM	7 transmembrane
A ₂₆₀ nm	Absorbance at 260nm
Ab	Antibody
Ara	Arabinose
BSA	Bovine serum albumin
cDNA	complementary DNA
cGMP	Cyclic guanosine monophosphate
CLOCK	Circadian locomotor output cycles kapu
CNG	Cyclic nucleotide gated
DEPC	Di-ethyl pyro-carbonate
DNA	Deoxyribonucleic acid
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide tri-phosphate
DTT	Dithiothreitol
E.coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid electrophoresis
EST	expressed sequence tag
FBS	Foetal bovine serum

FtsH	Filamentous temperature sensitive H1
gDNA	genomic DNA
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GPCR	G protein-coupled receptor
GTP	Guanosine triphosphate
GTP	Cyclic guanosine triphosphate
H ₂ O	Water
Hb	Haemoglobin
His ₈	Polyhistidine
hr	hour/s
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IPTG	Isopropylthio-β-D-galactoside
Kb	Kilobase
kDa	Kilo Dalton
LB	Lysogeny broth
MBS	Modified Barth's saline
MCR	Multiple cloning region
MCS	Multiple cloning site
Min	minute/s
MP	Membrane protein

MPOS	3-(N-morpholino) propanesulfonic acid
MR	Modified fog Ringer's
MS222	Tricaine methanesulfonate
MI	Micro-Litter
ng	Nanograms
NMR	Nuclear magnetic resonance
nt	Nucleotides
OD	Optical density
ORF	Open reading frame
PAGE	Poly-acrylamide gel electrophoresis
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PDE	Phosphodiesterase
PNP	Parapinopsin
PRC	Photoreceptor cell
P _{tet}	Tetracycline promoter
RNA	Ribonucleic acid
RNA	Ribonucleic acid
Rpm	Rounds per minute
SCN	Superchiasmatic nuclei
SD	Shine Dalgarno sequence

SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel
SSC	Saline-sodium citrate
SSP-PCR	Strand specific primer- polymerase chain reaction
TAE	Tris acetate Ethylenediaminetetraacetic acid
TE	Tris-HCl EDTA
TEMED	Tetramethylethylenediamine
TEV protease	Tobacco etch virus protease
UV	Ultra violet
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactoside

Chapter 1 Introduction

1.1 Structure of the Eye

The eye is like a camera. The external object is seen in the same way that the camera takes the picture of an object. Light enters the eye through a small hole called the pupil and is focused on the retina (Figure 1.1), which is like a camera film. The eye also has a focusing lens, which focuses images from different distances on the retina. The coloured ring of the eye, the iris, controls the amount of light entering the eye. It closes when light is bright and opens when light is dim. A tough white sheet called the sclera covers the outside of the eye. The front of this sheet is transparent (the cornea) in order to allow the light to enter the eye. Ciliary muscles in the ciliary body control the focusing of the lens automatically. The choroid layer forms the vascular layer of the eye, supplying nutrition to the eye structures. The image formed on the retina is transmitted to the brain by the optic nerve. The image is finally perceived by the brain. A jelly like substance called vitreous humor fills the space between the lens and the retina. The lens, iris and cornea are nourished by clear fluid, aqueous humor, formed by the ciliary body which fills the space between lens and cornea (Schulz *et al.*, 2004). The eye converts light signals into electrochemical signals in neurons by photoreceptor rod and cone cells. Sight or vision begins from the photoreceptors, which absorb light stimuli and transduce the signal to the nervous system. The function of the simplest photoreceptor cell in conscious vision is to connect light vision to movement. In higher organisms the optical system is very complex, and works by collecting and regulating light from the surroundings by a diaphragm, which regulates the light, and a lens, which focuses this regulated light to make an image which is converted into electrical signals. Optical nerve fibers transmit these electrical signals to the brain via complex neural pathways (Hubel, 1988).

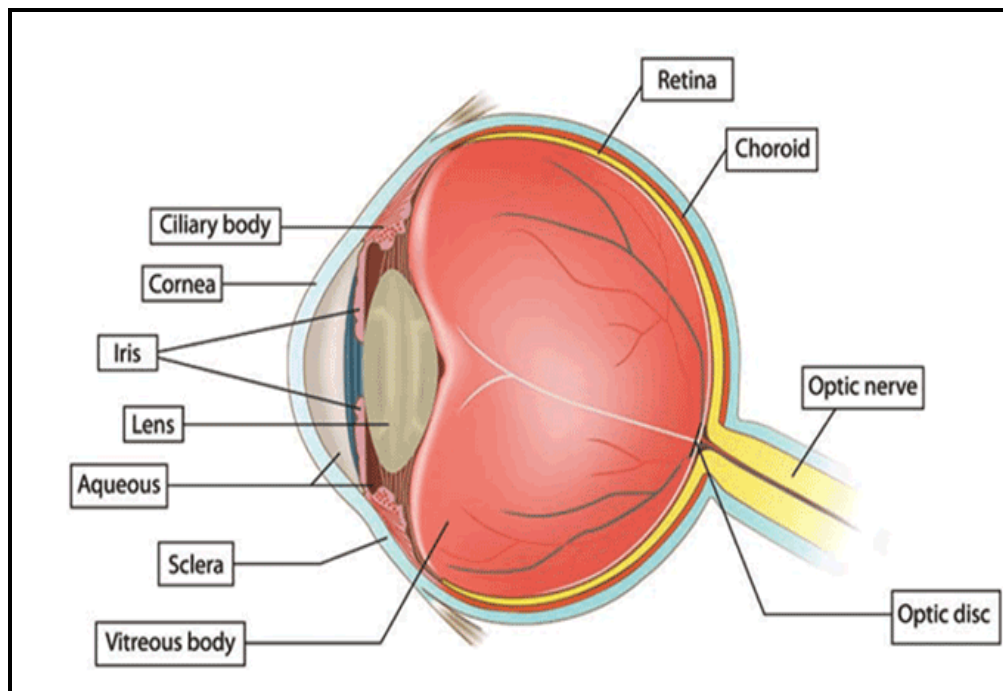


Figure1.1: Structure of the eye

The figure above shows the structure of an eye. The lens focuses the light onto the retina. The retina is similar to a camera's image sensor, composed of millions of photoreceptors called rods and cones. The visual information recorded by these photoreceptors is collected by the optic nerve.

(Figure adopted: <http://www.glaucoma.org/glaucoma/anatomy-of-the-eye.php>)

1.2 Structure of Retina

The retina lines the back of the eye. The optic nerve contains the ganglion cell, axons running to the brain and, additionally, incoming blood vessels open into the retina to vascularize the retinal layers and neurons. A section of a portion of the retina reveals that the ganglion cells (the output neurons of the retina) lie innermost in the retina closest to the lens and the photoreceptors (the rods and cones) lie outermost in the retina against the pigment epithelium and choroid (Figure 1.2). Light must, therefore, travel through the thickness of the retina before striking and activating the rods and cones. All vertebrate retinas are composed of the outer layer of the rods and cones and the inner layer of the horizontal cells, bipolar and the ganglion cells (Hubel, 1988).

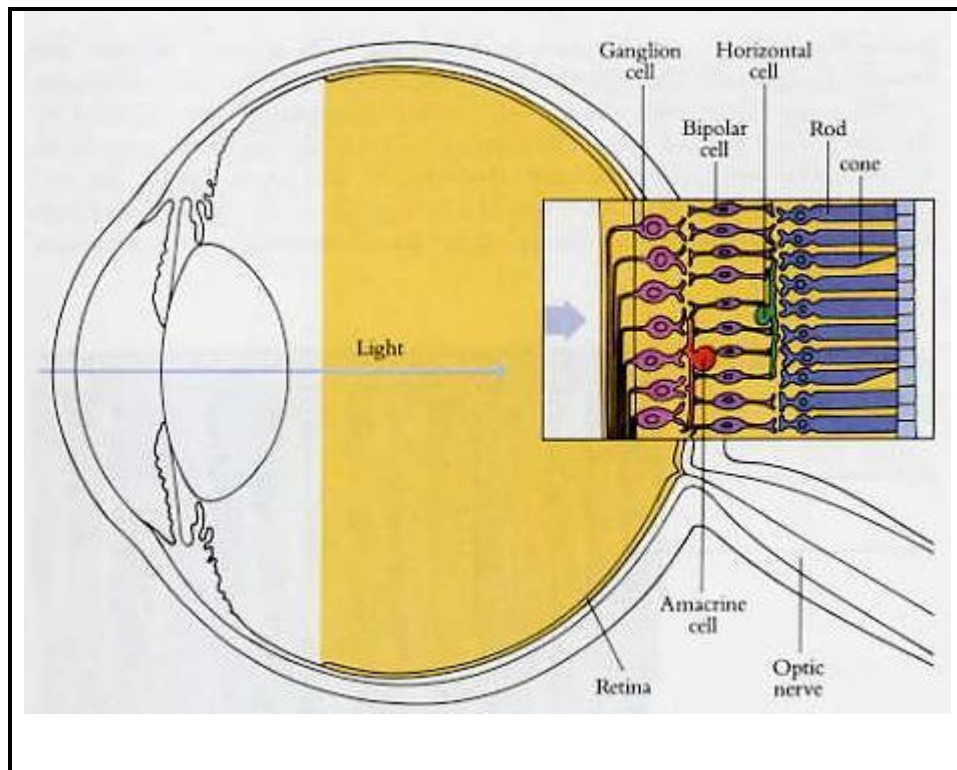


Figure1.2: Structure of the Retina

Section of the retina depicting overall arrangement of retinal layers. The arrow shows the direction of light coming from the vitreous humour and which is absorbed by pigment epithelium (Hubel, 1988)

1.3 Function of the eye

Eyes are very important organs in a body as we see the world through them. Eyes perform dual functions of “image detection” and clock measurement via “brightness detection”. Hence eyes are sense organs which connect light and vision and also light and time. Argentinian novelist Julio Cortazar wrote in 1965 that “time enters through eyes”. Other than rod and cone photoreceptors, which are involved in vision, most vertebrate species have non-image forming photoreceptors, sensitive to direct light, and transfer time information from the environment and act as circadian light receptors (Foster and Kreitzman, 2014).

1.4 Brain

The brain is the most complex organ in the body of all vertebrates and most invertebrates, positioned in the head, and the centre of the nervous system. 15-33 billions neurons are present in the highly complex vertebrate brain and communicate with each other by long fibres called axons through synapses (Hubel, 1988). Some non-visual

photoreceptors are also present in parts of the hindbrain called the pineal complex and parapineal organ that are involved in deep brain photoreception (Vigh *et al.*, 2002).

1.4.1.1 The Frog Brain

Most parts of the frog brain correspond with the human brain. The frog brain is divided into three main parts forebrain, midbrain and hindbrain described and is shown in Figure 1.3 below:

I- Forebrain

It consists of a pair of olfactory lobes; a pair of cerebral hemispheres, a pineal body, two optic lobes, a diencephalon. The pineal gland is present in the diencephalon between left and right hemispheres. Olfactory lobes are involved in the sense of smell.

II- Midbrain

A pair of optic lobes is present in the middle part of brain and is involved in integration of nerve impulses from the eyes. The large size of optic lobes indicate the importance of sight.

III- Hind brain

The hind brain comprises the cerebellum and medulla oblongata. The cerebellum coordinates the body movements. The functions performed by the medulla oblongata are, regulation of respiration, digestion and other autonomic functions. The foramen magnum is the point where the medulla oblongata connects with the spinal cord (Wikzynski, 2009).

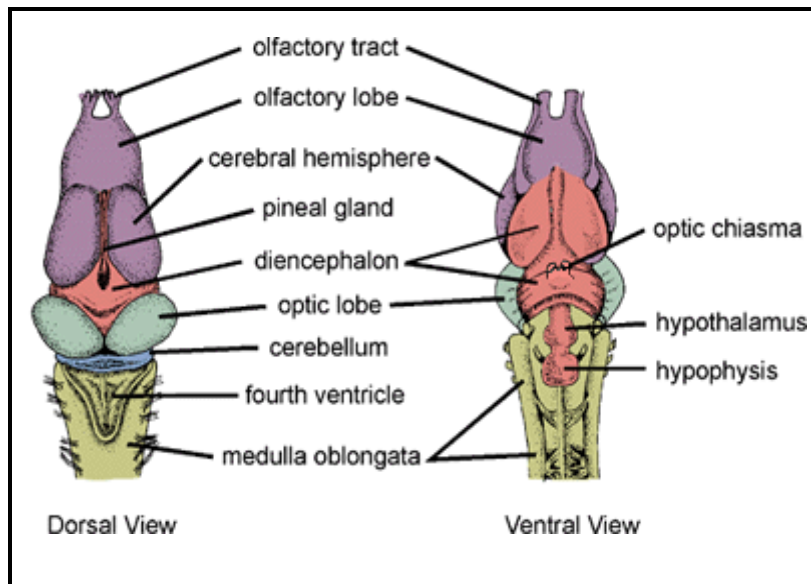


Figure 1.3 Frog's brain dorsal and ventral view

Figure above shows the frog's brain dorsal and ventral view. The pineal gland is present between the cerebral hemispheres, extends into diencephalon and is clearly visible on the dorsal side view of the brain.

(Figure adopted: <http://www.tutorvista.com/biology/nervous-system-of-a-frog>)

1.5 The Vertebrate Pineal organ

1.5.1 Pineal and parapineal organs of Lamprey and Teleost

Among vertebrates, the structure of lamprey's pineal organ is the simplest one and is composed of a small flattened vesicle attached to the skull, divided into pineal and parapineal organs separated by an atrium (Figure 1.4). The pineal body has a dorsal and ventral wall. Photoreceptors are present only in the ventral part of the pineal and parapineal organ. This can be revealed due to the binding of retinal rhodopsin antisera to the pineal organ. A small round body, called the pineal organ, located on left side of the brain above the epithalamus in teleosts also shows the presence of a visual rhodopsin like pigment (Vigh *et al.*, 2002).

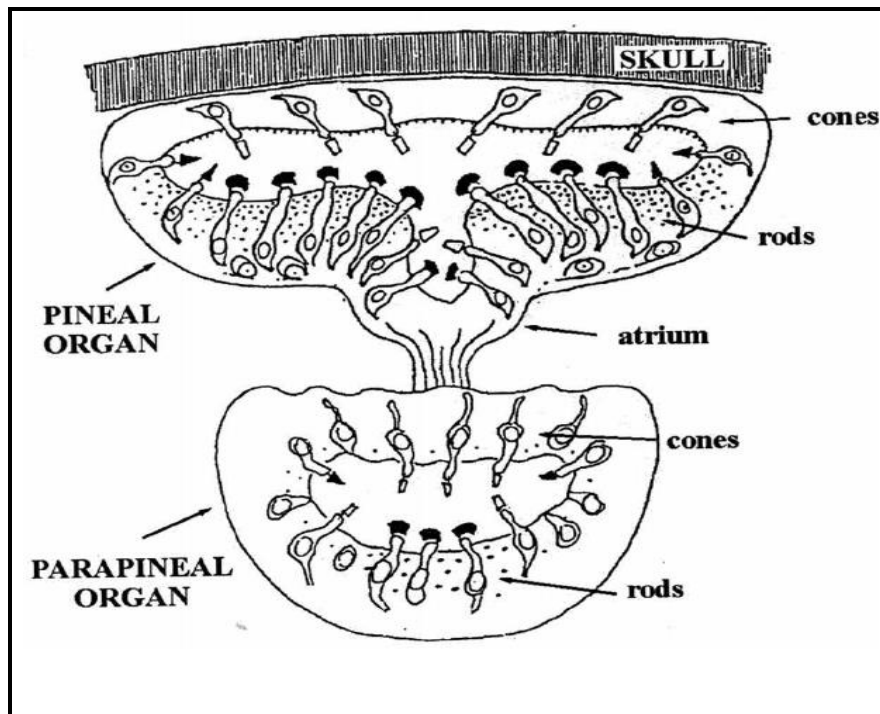


Figure 1.4 Diagrammatic representation of the Lamprey pineal organ

The above figure shows the lamprey pineal and parapineal organs inside the skull, with rod and cone cells. The pineal organ continues towards the parapineal organ and dilates into a small atrium. The rod and cone cells are only present on the ventral walls of the organs (Vigh *et al.*, 2002).

1.5.2 The Pineal Complex of Amphibians

The pineal complex of amphibians is composed of the pineal organ and frontal eye. In newts and salamanders the pineal complex is represented by the pineal organ only (Takahama, 1993), while in frogs the pineal complex is made up of the pineal organ, parapineal complex and frontal eye, shown in Figures 1.5 (a & b) and 1.6.

1.5.2.1 Frontal eye

The frog's extracranial frontal eye is buried under the skin between the lateral eyes. The extracranial frontal eye can sense direct solar light and act in positive or negative phototaxis. The cone type photoreceptors are present in the frontal eye (Vigh-Teichmann and Vigh, 1990).

1.5.2.2 The Pineal organ

The pineal organ, also known as an intracranial pineal organ, was thought to be described for the first time ~300 BC ago by Herophilus (Ung and Molteno, 2004). It is

present in almost all non-mammalian organisms as the main source of extra-ocular photoreceptors (Foster and Soni, 1998). Pineal secretions control the function of the adeno- and neurohypophysis, the thyroid and parathyroid, adrenal cortex and medulla, endocrine pancreas and gonads (Ung and Molteno, 2004). Immunocytochemistry with various antisera shows the presence of rods and cones in the pineal organ of *Rana* species. The size of the pineal organ at larval stages is relatively large, and tadpoles use these photoreceptors underwater (Vigh *et al.*, 2002).

1.5.2.3 The Parapineal organ

The pineal complex is composed of a midline pineal organ and a left sided parapineal organ. Nuclei called “bilateral habenulae” present in the diencephalon, form part of the forebrain. These nuclei along with parapineal organ are the primary circadian rhythm pacemaker, and convey information from forebrain to midbrain (Snelson *et al.*, 2008). In higher vertebrates SCN (superchiasmatic nuclei) are the major pace maker (Bianco and Wilson, 2009). This thesis is based on “Parapinopsin”, a novel UV sensitive pigment protein which shows expression within the parapineal organ (Blackshaw and Snyder, 1997). Figure 1.6 shows the parapineal organ of the frog.

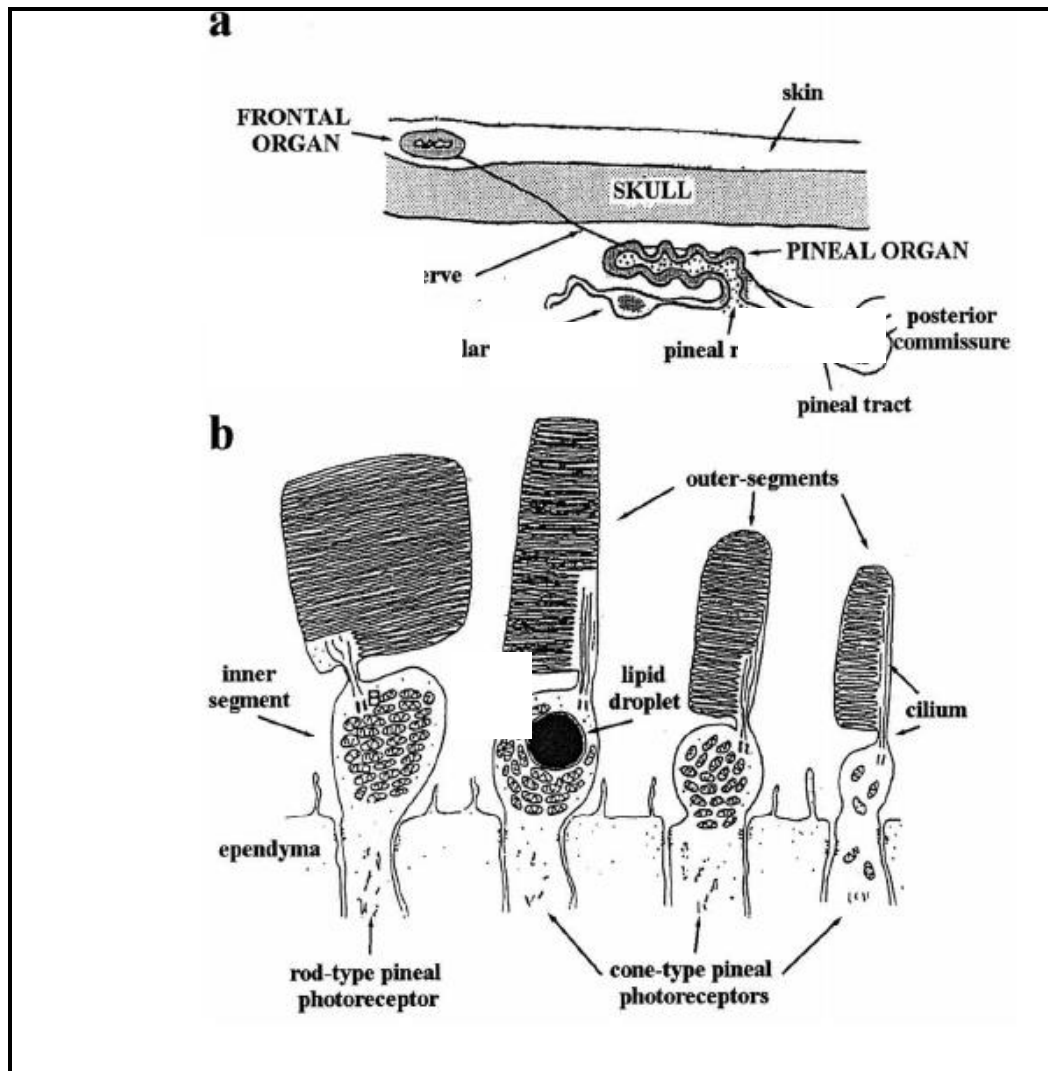


Figure 1.5 Diagrammatic presentation of the Pineal complex of the Frog

The part "a" of the above figure shows the extracranial frontal organ of the frog present under the skin and connected to the intracranial pineal organ via frontal nerve, and the pineal organ connected to the brain through the pineal tract. Part "b" of the above figure shows the photoreceptors present in the pineal organ (Vigh et al., 2002)

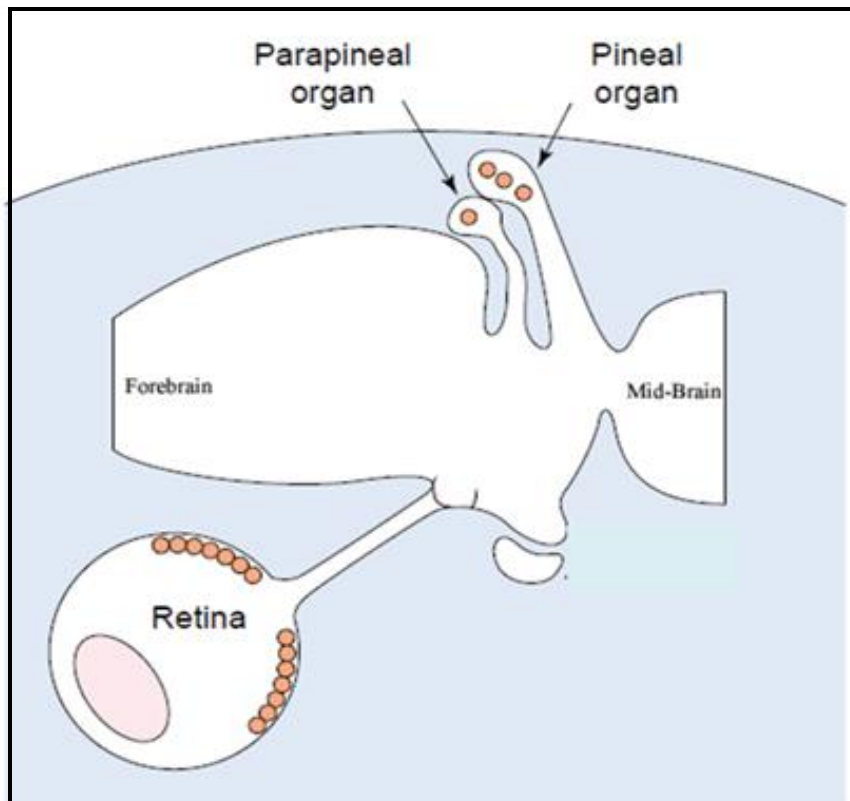


Figure1.6 Diagrammatic presentation of the Frog's pineal, parapineal organs and the retina (Soni and Foster, 1998)

The figure above represents the position of the pineal and parapineal organs in the frog's brain. The orange circles show the presence of rod and cone photoreceptors in the retina, pineal and parapineal organs.

1.5.3 The Pineal Organ in Reptiles

Reptiles have an intracranial pineal or extra cranial parietal eye, except the alligator which has no pineal (Tosini *et al.*, 2001) only a dorsal lens and ventral retina containing photoreceptor cells. The pineal organ is more complex when compared to the amphibian pineal organ and more lobulated when compared to the parietal eye. The pineal organ is attached to the inner side of the skull and connected to the epithalamus through a narrow stalk. The parietal eye contains photoreceptor cells, neurons and glial elements (Vigh *et al.*, 2002). In *Anolis carolinensis* the photoreceptors do not respond to anti-rhodopsin antisera and anti-chicken iodo-opsin, but three visual opsin and p-opsin genes were found to express in the parietal eye and pineal organ (Debrececi *et al.*, 1998). In lizards, the parietal eye and the pineal organ are involved in maintaining several body functions in response to light, indicating the regulation of body temperature (Engbretson and Hutchison, 1976). In all species, the pineal outer segment is more differentiated in young

animals as compared to adult ones depicting the importance of environmental light interaction during post-hatching stage with internal body rhythm (Fejer *et al.*, 2001)

1.5.4 The Mammalian Pineal gland

The mammalian pineal organ is called the pineal gland, and its cells are called pinealocyte which function to secrete a hormone “melatonin”. Its production is inhibited by the presence of light, perceived by photoreceptors in the retina, and mediated via the suprachiasmatic nucleus (SCN) and sympathetic fibres. Hence pinealocytes are not sensitive to direct light (Stehle *et al.*, 2001). Experimental studies reveal that several phototransduction cascade molecules are present in the mammalian pinealocyte and that the photoreception mechanism depends upon the intact retina (Korf *et al.*, 1998). In mammals the pinealocyte phototransduction cascade photosensitivity decreases with age. Human pinealocytes also preserve photoreceptor properties e.g. rod and cone opsins. In humans, like other mammals, light also penetrates the skull (Vigh *et al.*, 2002).

1.6 Circadian Rhythm

All living organisms have a special relationship to light in order to regulate their life processes, particularly vertebrates which depend largely on light to synchronize their physiology (Provencio *et al.*, 2000). Inside the body the biological process's rhythmical oscillations, repeating nearly 24 hours, are called “circadian rhythm” and are driven by a “circadian clock”. They are central biochemical mechanisms coordinated with day and night cycles which influence almost all aspects of life (Zelinski *et al.*, 2014). These endogenous biological clocks are affected by zeitgeber, exogenous signals, e.g. light, temperature and food. The endogenous circadian cycles are ubiquitous in nature and exist in plants, animals, fungi, algae and bacteria, and anticipate / synchronize life on earth for changes in light and dark. An example of external and internal time desynchrony is Jet-Lag (Foster and Kreitzman, 2014). Circadian rhythm is represented in cyclic form in Figure 1.7 (a).

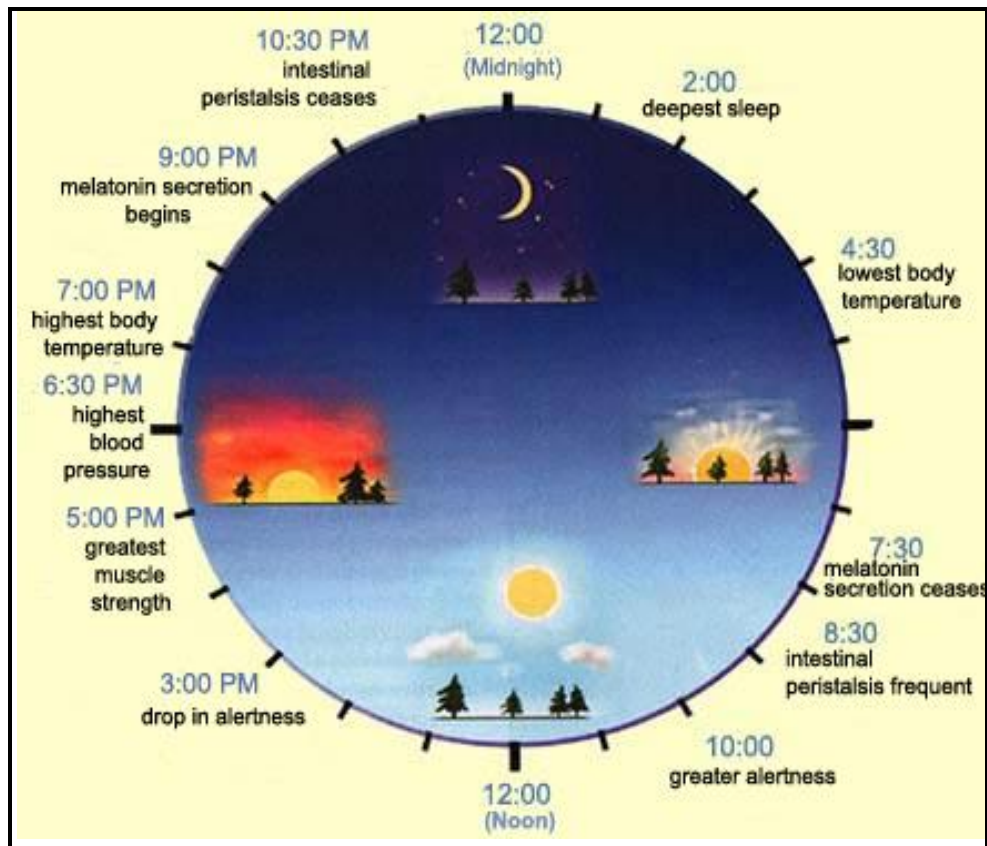


Figure1.7 (a) Cyclic overview of circadian rhythm

The above figure represents the circadian rhythm of day and night 24 hours cycle and the physiological changes that the biological clock brings inside the body.

(http://thebrain.mcgill.ca/flash/i/i_11/i_11_p/i_11_p_hor/i_11_p_hor.html)

The clock system is highly plastic and adjusts itself to external cues to reset itself (Masri *et al.*, 2013). The endogenous circadian clock, in accordance with external cues, is controlled by the pineal organ via the SCN (suprachiasmatic nuclei in brain) of higher vertebrates e.g. humans (Refinetti and Menaker, 1992). Due to this rhythm, animals survive in changing environments (Vallone *et al.*, 2005). An overview of the mechanism is shown in Figure 1.7 (b)

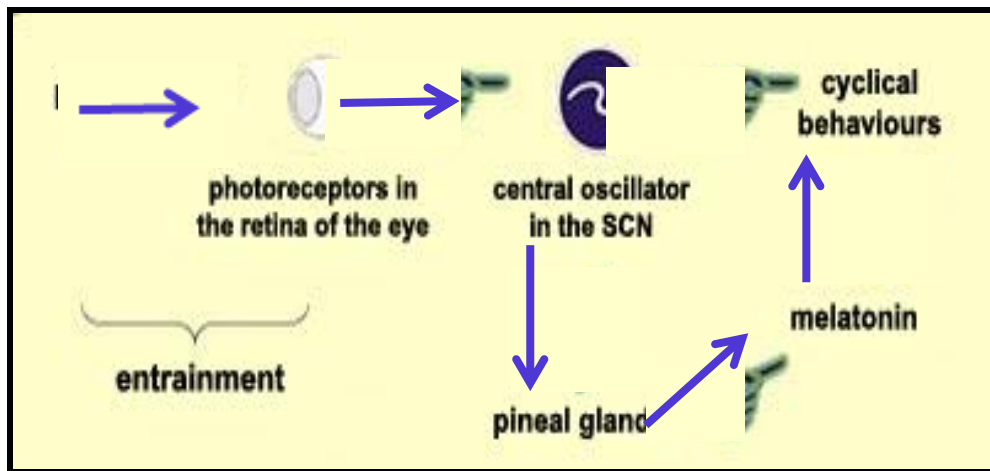


Figure 1.7 (b) Schematic representations of steps involved in melatonin production

Figure above shows how the circadian retinal photoreceptor activation by light results in melatonin secretions from the pineal gland, which is under the master control of the suprachiasmatic nuclei (SCN).

(http://thebrain.mcgill.ca/flash/i/i_11/i_11_cr/i_11_cr_hor/i_11_cr_hor.html)

1.6.1 Circadian Rhythm in Vertebrates

In vertebrates, both the pineal gland and the retina are sites of circadian rhythm allowing communication to the rest of the body via melatonin (N-acetyl-5-methoxytryptamine) in response to changes in day and night and seasonal changes. Steps involved in melatonin synthesis are shown in Figure 1.8. Melatonin concentration is higher at night and lower in day time. Hence melatonin production is dependent on zeitgebers e.g. presence and absence of light (Takahashi *et al.*, 2001). In most species the pineal is the main site for melatonin synthesis and has high level of melatonin, except frogs where the level of melatonin is higher in the retina suggesting the neuroendocrine function of retina as well (Isorna *et al.*, 2006).

Melatonin functioning in amphibians controls as circadian rhythmical lightening of skin color, inhibitory and excitatory effects on reproductive system, changes to metamorphosis (Filadelfi, Castrucci, 1996) and acts as a powerful antioxidant to scavenge free radicals (Poeggeler *et al.*, 1994).

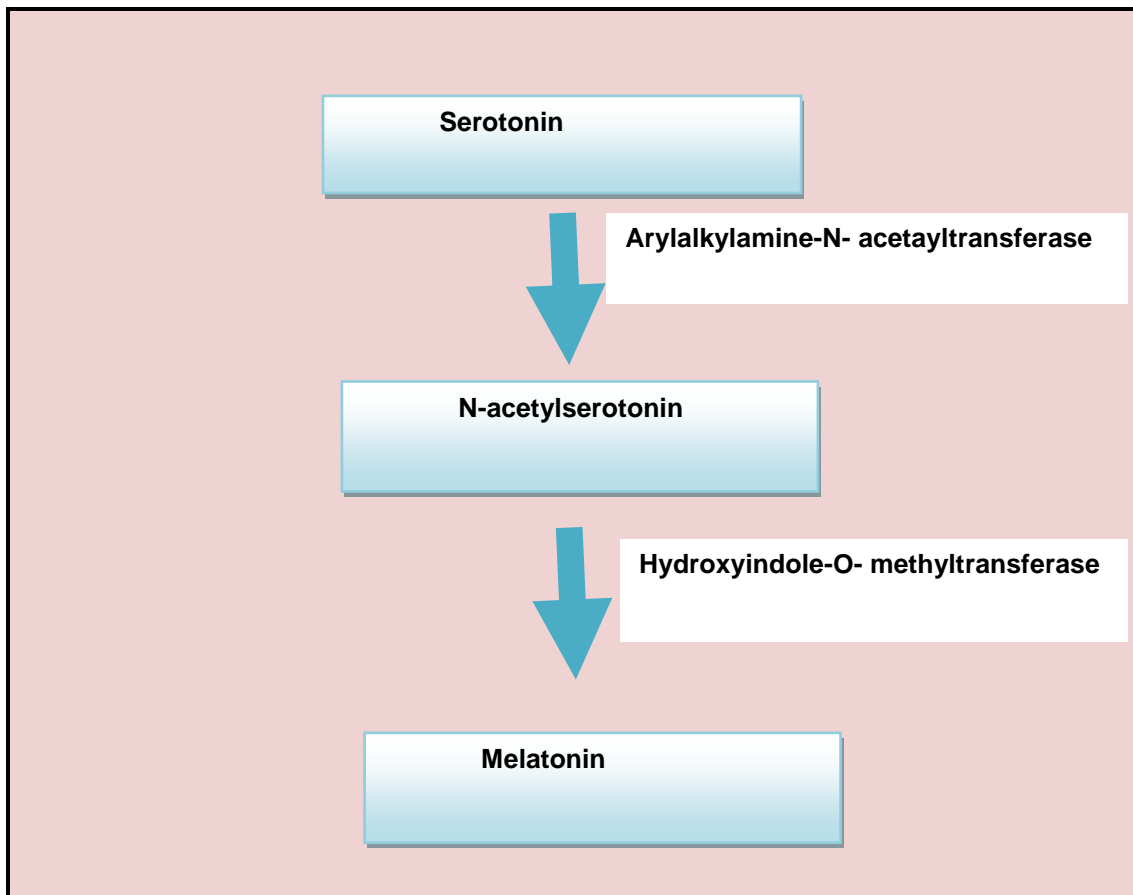


Figure1.8 Steps involved in vertebrate melatonin synthesis (Isornaa *et al.*, 2006)

The figure above shows the path of melatonin synthesis from serotonin in vertebrate retina and pineal.

1.6.2 Molecular Circadian Clock

Several circadian genes in the molecular clock control environmental cycles, circadian rhythm. The molecular control of these genes is present in retinal specific photoreceptor cells that send signals to the circadian master control regulator, suprachiasmatic nucleus "SCN", which results in the synthesis of melatonin in the pineal gland. The SCN is the primary circadian pacemaker in many vertebrates, but is not well characterized in frogs (Provencio *et al.*, 1998). These genes work in a feedback loop (Kelleher *et al.*, 2013). The transcription/translational feedback mechanism is shown in Figure 1.9. Transcription factors "CLOCK" (Circadian Locomotor Output Cycles Kapu) and "BMal" (Brain-muscle Arnt-like) and Npas2 (Neuronal PAS domain-containing protein 2) in the nucleus binds with circadian enhancer sequence CACGTG-type "E-box" and results in the

transcriptional activation of cryptochromes (Cry1-2) and Period genes (PER), which are related to repair of UV based DNA damage. “Cry1-2” and “PER” act as negative feedback genes and have been mainly studied in vertebrates. Two cryptochromes “Cry1-2” genes were found in *Xenopus tropicalis* ovary (Kubo *et al.*, 2010). CLOCK, cryptochrome, Bmal1 and Period genes were successfully cloned from *Xenopus laevis* (Green, 2003).

The circadian rhythm is controlled by the “retinal clock” in *X. laevis* due to retinal expression of different homologues of cryptochrome genes, although Cry expression was also found in brain, heart, liver, spleen and testis of *X. laevis*. The presence of different homologues of Cry gene in retina may represent more than one function of the Cry gene in *X. laevis*. The retinal clock of *X. laevis* can be reset by direct exposure to light. In higher vertebrates the SCN clock can be reset indirectly via signalling through the eye (Zhu and Green, 2001).

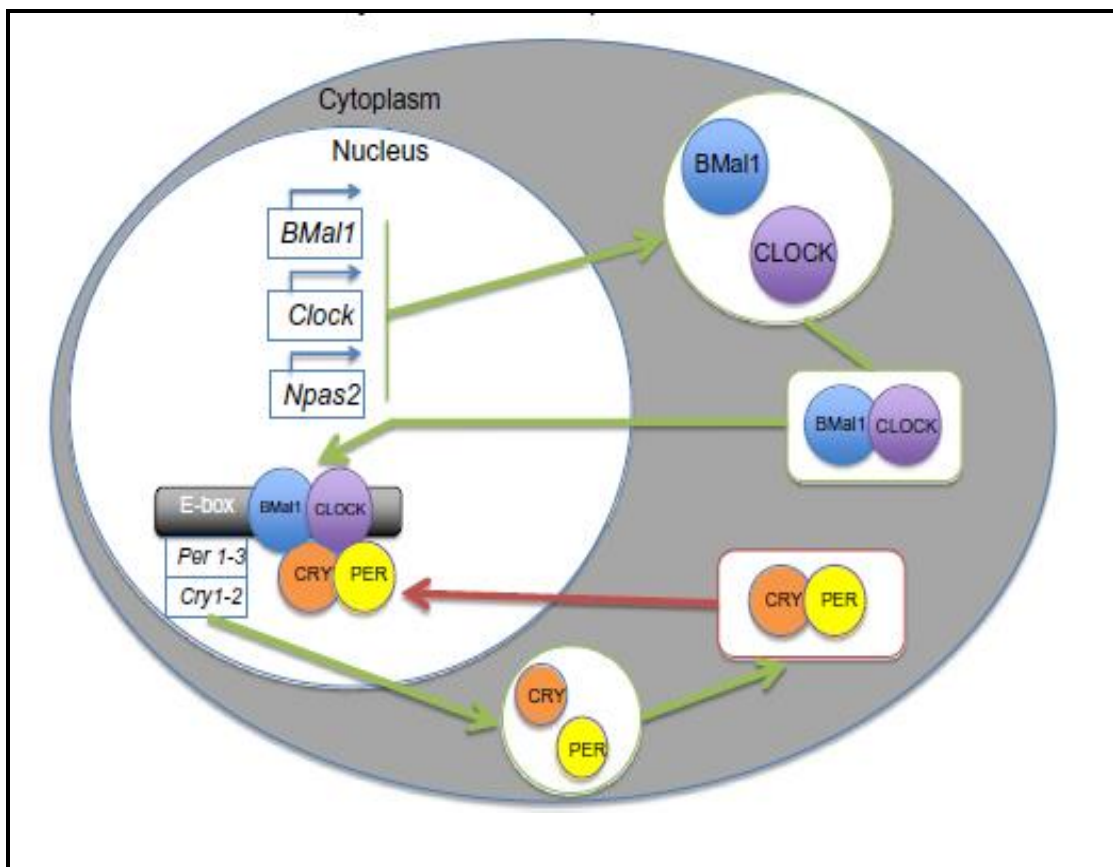


Figure1.9 Feedback mechanism of circadian clock (Kelleher *et al.*, 2014)

Above figure shows the overall genes involved in vertebrates to control the feedback loop of the circadian clock. The molecular clock rhythmically transcribes

circadian genes in a feedback loop manner, and is followed by degradation of protein complexes that negatively impact on their own production. The primary feedback loop is illustrated in this figure. In the nucleus three helix loop helix domains are found in transcription factors from Clock, BMal1 and Npas2 genes which form heterodimers (e.g. BAL1/CLOCK) and result in activation of transcription of the period genes (Per1 and Per2) and cryptochrome genes (Cry1 and Cry2) by binding to the E-box promoter. The PER/CRY protein complex represses its own gene transcription by blocking BMAL1/CLOCK activity.

1.7 Photosensory Organs

Organisms respond to changes in light using different photosensory organs. These photosensitive organs include primarily eyes, but some photosensory cells are located in the pineal organ, brain or in peripheral tissues (Forsell *et al.*, 2001). There are two types of photoreceptors according to their function:

Visual photoreceptors

Non-visual photoreceptors

1.7.1 Visual Photoreceptors

All vertebrates contains visual photoreceptors present in the “lateral” eyes i.e. the retina also called “ocular visual photoreceptors” e.g. rods and cone cells (shown Figure1.2), involved in visual process (Ung and Molteno, 2004).

1.7.2 Non-visual Photoreceptors

Non-visual photoreceptors are positioned in the “lateral” eye retina as well as in the deep brain i.e. pineal and parapineal organ or frontal or parietal eye. These photoreceptors are involved in synchronization of light based organism’s endogenous circadian clock (Vigh *et al.*, 2002, Foster and Kreitzman, 2014). Extra ocular photoreceptors are the main photoreceptors in non-mammalian vertebrates or in lower vertebrates, and are present in the pineal and parapineal organs of the mid brain (Foster and Soni, 1998).

1.8 Classification of Photoreceptors

The animal photoreceptor cells (PRCs) are classified into two main groups: ciliary photoreceptors and rhabdomeric photoreceptors (Koyanagi *et al.*, 2005). PRCs have large membrane surface area for the storage of a photopigment called opsin (Arendt *et*

al., 2004). Both types of PRCs are present in vertebrates and invertebrates, strongly suggesting that our common ancestor used both PRCs (Shichida & Matsuyama, 2009).

1.8.1 Ciliary Photoreceptors Cells

Their photopigment is borne on the disc membrane of the modified cilium (Koyanagi *et al.*, 2005, Koyanagi and Terakita, 2013), shown in Figure 1.10. The Ciliary photoreceptors constitute a large class of photoreceptor cells including rod, cone cells, and photoreceptor cells of the light sensitive pineal organ (Arendt *et al.*, 2004).

1.8.2 Rhabdomeric Photoreceptors Cells

In Rhabdomeric photoreceptors cells, shown in Figure 1.10, the apical cell surface is invaginated into several folds, and photopigment is borne on apical microvilli (Koyanagi and Terakita, 2013). These are the important photoreceptors cells involved in invertebrate visual processing (Arendt *et al.*, 2004). It is supposed that vertebrates light sensitive circadian photoreceptor cells are a modified form of the invertebrate's rhabdomeric photoreceptor cells.

1.8.3 Vertebrates Ciliary Opsins

Rod and cone opsins are examples of Vertebrates Ciliary opsins, and studies show that rod opsins evolved from four cone opsins (Yau and Hardie, 2009, Yokoyama 2000, Arendt & Wittbrodt 2001, Terakita 2005, Suga *et al.*, 2008, Shichida & Matsuyama 2009 and Lamb 2009).

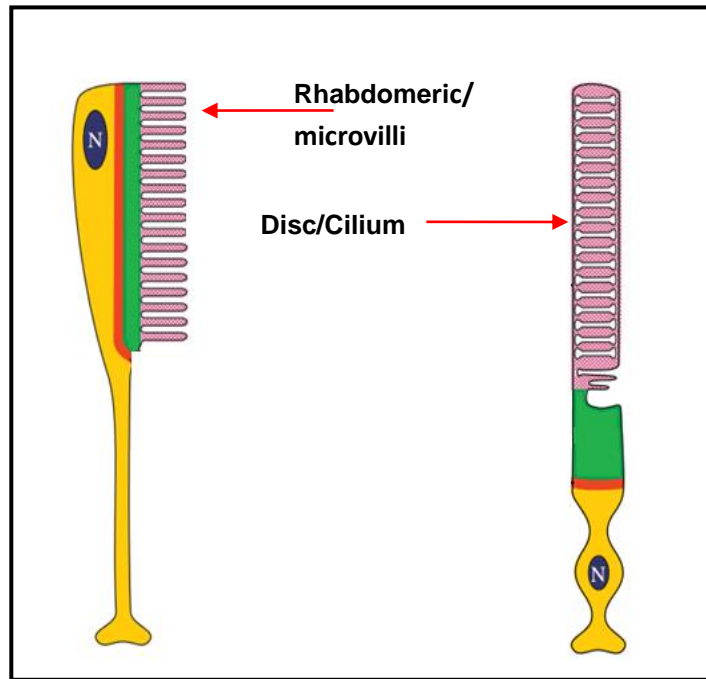


Figure 1.10 Diagrams of rhabdomeric and ciliary photoreceptor cells
(Lamb, 2009)

The figure above shows the structures of a rhabdomeric photoreceptor cell (*Drosophila*), and a ciliary photoreceptor cell (mammal), where N is the nucleus of the cell.

1.9 GPCR: The Opsins

G-protein coupled receptor (GPCR) signal transduction is very important for vertebrate physiology. GPCR act as receptors for various ligands like hormones, neurotransmitters, ions and photons *etc.* GPCR link the cell's environment with the external environment (Rosenbaum *et al.*, 2009). GPCR make the largest family of proteins with nearly 800 GPCR known in humans. The alternative name for GPCRs is 7TM (7 transmembrane) (Choe *et al.*, 2010).

The opsins belong to family "A" of the G protein coupled receptors (GPCR); making the largest group of 7 transmembrane receptors. Rhodopsin is the best characterized member of GPCR and can be used as a template in homology studies to standardize other GPCR (Shichida and Matsuyama, 2009; Lagerstrom and Schioth, 2008).

The first opsin sequenced in 1982 was “bovine rhodopsin”, and to date more than 1000 opsins have been sequenced and sequences are available online, ranging from Jellyfish to humans. Some opsins are involved in visual reception, others are involved in non-visual functions, the “circadian rhythm”. The function of most of the opsins is not known. The opsins have been classified into 6 sub-families with less than 20% sequence homology within the sub-family (Shichida and Matsuyama, 2009; Shichida and Yamashita, 2003; Terakita, 2005; Shichida and Matsuyama, 2009).

- i. The Vertebrate opsin/encephalopsin subfamily (vertebrate visual and non-visual pigments)
- ii. The Go opsin subfamily (molluscs and chordate amphioxus, opsin)
- iii. The Gs opsin subfamily (cnidarians opsin *i.e.*, jellyfish, coral)
- iv. The invertebrate Gq opsin subfamily (invertebrate visual pigment and melanopsin)
- v. The photoisomerase subfamily and the neuropsin subfamily

The Vertebrate opsin/encephalopsin, Go opsin and Gs opsin are expressed in ciliary photoreceptors. The Gq opsins are expressed in rhabdomeric photoreceptor cells (Shichida and Yamashita, 2003; Koyanagi and Terakita, 2013).

1.9.1 The Vertebrate Opsin/encephalopsin Subfamily

This group comprises vertebrate opsins and encephalopsins.

1.9.1.1 Vertebrate opsins

This group consists of vertebrates visual and non-visual opsins.

In 1992 Okano *et al.*, divided the vertebrate visual opsins into 5 groups shown in table

1.1 according to their functions.

Name of group	Type of opsin	Function
S group	Cone opsin	Absorbs ultra violet or violet light
M1 group	Cone opsin	Absorbs blue light
M2 group	Cone opsin	Green light
L group	Cone opsin	Absorbs red or green light
Rh group	Rod opsin	Absorbs blue/green light

Table 1.1 Classification and functions of vertebrate opsins (Shichida and Matsuyama, 2009)

The Table above explains the types of vertebrate visual opsins and their functions. The first four groups S,M1,M2 and L belong to cone opsin and the last, Rh group, belongs to rod opsin.

The expression of vertebrate visual opsin in non- visual photoreceptor cells like pineal organ has been found by Wada *et al.*, 1998; Mano *et al.*, 1999.

1.9.1.2. The Vertebrate Non-Visual Opsin

All vertebrate non-visual opsins have 40% sequence homology with vertebrate visual opsins (Shichida and Matsuyama, 2009). Non-visual photoreception is very important for living organisms such as the synchronization of the biological clock with the daily light–dark cycle (Shichida, 2003).

Non-visual opsin	Identification	Occurrence	Reference
Pinopsin	The first chicken retinal non-visual photoreceptor was identified in 1944. Its function speculated in circadian regulation. Exact function is unknown	Also found in reptile and amphibian pineal. Absent in teleost and mammals	Shichida, 2003; Okano <i>et al.</i> , 1994; Max <i>et al.</i> , 1995, Taniguchi <i>et al.</i> , 2001
Parapinopsin	Found in photosensitive pineal, parapineal organ and parietal eye. First identified in cat fish (discussed in more detail below)	Also found in teleost, , amphibian, green iguana	Blackshaw and Snyder 1997; Koyanagi <i>et al.</i> , 2004, Wada <i>et al.</i> , 2012
VA opsins	Vertebrate ancient opsin named after their early divergence in vertebrate opsin evolution, found in inner retina and brain.	Found in salmon	Soni and Foster 1997; Sato <i>et al.</i> , 2011
Parietiopsin	This is the opsin which is expressed in a photoreceptive organ called “third eye”	Expressed in lizards. Shows 40% sequence similarity with parapinopsin and VA opsin	Su <i>et al.</i> , 2006
VAL opsin	Isoform of VA opsin, studies reveal two copies of VAL which show differential expression, and may be involved in different roles	Present in teleost lineage	Kojima <i>et al.</i> , 2008; Sato <i>et al.</i> , 2011

Melanopsin	It is believed that		
	Melanopsin is expressed		
	Expressed in <i>Xenopus Laevis</i> dermal melanophores, iris and hypothalamic part of brain	in cells	Mitsumasa Koyanagi
		derived from an ancestral	2005, Arendt 2003,
		rhabdomeric	Koyanagi and Terakita
		photoreceptor	2013
		cell, and is found in retinal ganglion cells of mouse.	

Table 1.2 Vertebrates non-visual opsins

The Table above shows the different types of vertebrate non-visual opsin in column 1. The 2nd and 3rd columns give information about the name of the animal in which the protein was first identified, and related information. In column 4, the references are given.

1.9.2 Encephalopsins

These opsins were named encephalopsins after their first identification in mammalian brain and also in testis (Blackshaw & Snyder 1999). Encephalopsin expression also has been found in liver, kidney, pancreas, and heart muscle as well as in retina (Halford *et al.*, 2001). These opsins are also called “Panopsin” due to their broader range expression in different tissues (Shichida & Matsuyama, 2009).

1.9.3 Go Opsins

Go opsin expression has been reported in molluscs and chordate amphioxus only. Go opsins, like vertebrate opsins; are expressed in ciliary photoreceptor cells. (Kojima *et al.* 1997; Koyanagi *et al.*, 2002).

1.9.4 Gs Opsins

Gs opsins are found in cnidarians e.g. jelly fish, sea anemone, hydra and corals. These opsins are also expressed in ciliary photoreceptor cells (Plachetzki *et al.*, 2007; Suga *et al.*, 2008).

1.9.5 Gq Opsins

Gq opsins belong to arthropods and molluscs visual opsin as well as melanopsin. Melanopsins are present in vertebrates, but they are in close resemblance with invertebrate's visual opsins (Briscoe & Chittka 2001; Koyanagi *et al.*, 2008; Provencio *et al.*, 2000). As discussed above, melanopsin are photosensitive opsins involved in

circadian rhythms in vertebrates and higher mammals (Hattar *et al.*, 2003; Lucas *et al.*, 2003; Panda *et al.*, 2003).

1.9.4 Photoisomerases

The photoisomerase family comprises retinochrome, RGR opsins and peropsins. Photoisomerases bind only to all-trans retinal, and produce 11-cis retinal (Radu *et al.*, 2008; Koyanagi *et al.*, 2002; Tarttelin *et al.*, 2003). This will be discussed in detail later.

The phylogenetic relationship of opsin is shown in Figure 1.11.

1.10 Molecular Machinery Involves in Light Sensitive Visual System

The visual system components of all living organisms from jelly fish to humans share common features, and the ultimate goal of the visual system machinery is to see. Light sensing by rhodopsin is the first step in vision. Rhodopsin absorbs light that induces structural changes in its structure and activates downstream molecular machinery. The “G” protein triggers the enzymatic cascade and finally produces an “electrical response” in the photoreceptor cells (Kristiansen, 2004). The amplification of signalling depends upon rhodopsin which activates a number of G proteins. This in turn activates different “G-sub type” proteins which act through different signalling pathways (Shichida & Matsuyama, 2009; Koyanagi and Terakita, 2013).

1.10.1 Light Switch Molecules

Latest research reveals the importance of opsin-based pigments due to the “light switch” properties of these molecules, which can be used for target cells to induce particular physiological responses by exogenous stimuli using light instead of chemicals (Koyanagi and Terakita, 2013).

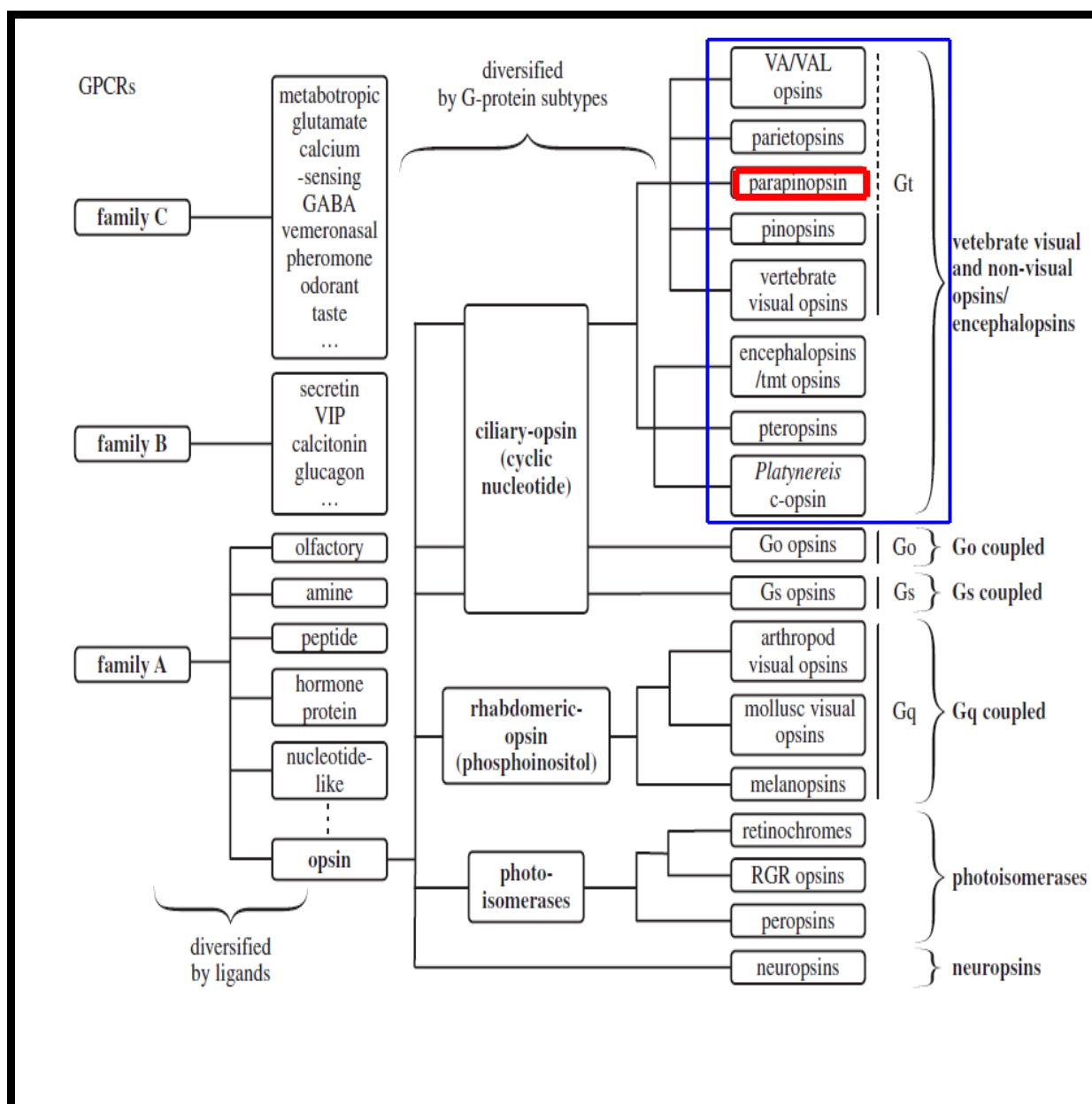


Figure 1.11 Schematic representation of the opsins phylogenetic relationship (Shichida and Matsuyama, 2009)

Scheme above shows the overall division of GPCRs into three families A, B and C. Opsins belong to family “A” of GPCR which is divided into ciliary opsin and rhabdomic opsin and photoisomerases. The Ciliary opsins expressed in ciliary photoreceptor cells are characterized by cyclic nucleotide signalling while rhabdomic opsin expression in rhabdomic photoreceptor cells is characterized by the phosphoinositol signalling cascade. The photoisomerases are characterized by stereospecific photoisomerases. The region of interest, the vertebrate visual and non-visual opsins are highlighted inside a blue box, and Parapinopsin is shown inside the red box which will be discussed below.

1.10.2 Rhodopsin

Rhodopsin, also called opsin-based-pigment (Koyanagi and Terakita, 2013), is a seven pass transmembrane helical protein consisting of an apoprotein part called “opsin” and a prosthetic group, “chromophore”. The N terminus of the opsin protein is in the lumen of the disc membrane and the C terminus is towards the cytoplasmic side. The chromophore is covalently bound with opsin through a Schiff base linkage to a lysine residue in helix number 7 (Figure 1.12). This Schiff base linkage can be protonated or deprotonated depending upon the surrounding environment. The protonated Schiff base results in delocalization of π electrons resulting in absorbance of visible light. The chromophore is protonated when it is bound with rhodopsin resulting in absorbance of visible light (Koyanagi and Terakita, 2013). Hence opsins provide a light absorbing environment for different wavelengths and colours and also an interface to bind with G proteins (Shichida & Matsuyama, 2009).

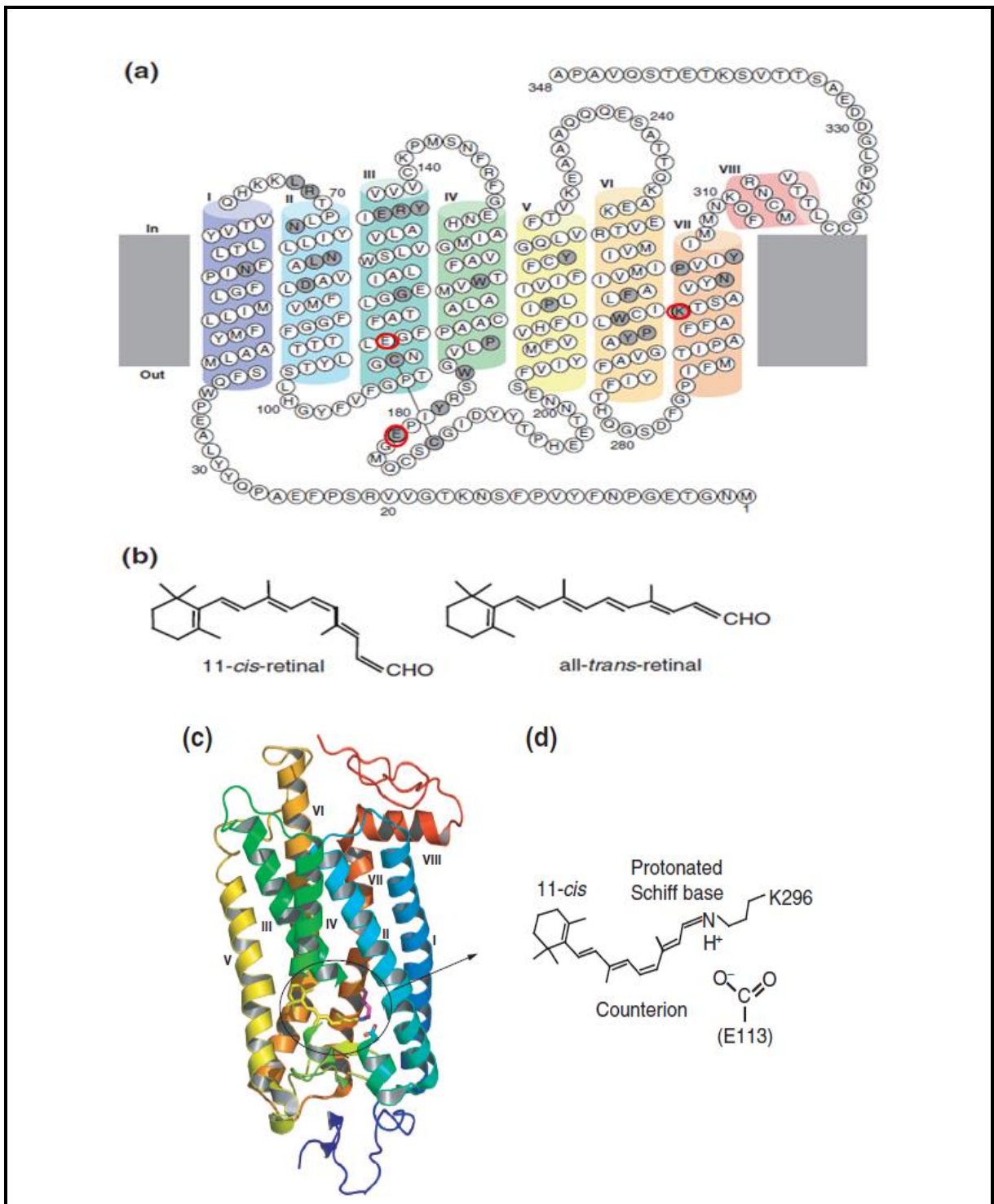


Figure 1.12 Bovine Rhodopsin Structure (Terakita, 2005)

Part (a) of above figure shows the conserved amino acids with grey background in the secondary structure of rhodopsin protein where the counterion ion and retinal binding site amino acids are highlighted with red circles at positions E113, and K296 respectively. (b) The structures of all-cis retinal and all-trans retinal are shown. (c and d) seven helices (HI-HVII) and the VIII helix, at the intracellular side and in line with membrane. Bound chromophore, retinal, is shown (sphere) and the close up structure of rhodopsin chromophore is shown with important amino acids positions.

1.10.2.1 The Chromophore

The chromophore is retinaldehyde (vitamin A) an important component that senses light.

There are four types of retinaldehydes:

- The Retinal (A1)
- The 3,4-dehydroretinal (A2)
- The 3-hydroxyretinal (A3)
- The 4-hydroxyretinal (A4)

Both vertebrates and invertebrates most frequently use “A1 retinal” as a chromophore.

Retinal has several isomeric forms like all-trans, 13-cis, 11-cis and 9-cis. In the dark state the opsin mainly binds with 11-cis retinal isomeric form of the chromophore. In the light the chromophore changes its conformation, and opsin binds the all trans form of retinal (Pope *et al.*, 2013; Shichida and Matsuyama, 2009).

1.10.3 Classification of Animal Phototransduction Cascade

Two distinct groups of animal Phototransduction are found, classified on the basis of evolutionary and functional similarity (Koyanagi *et al.*, 2008).

- Phosphoinositol signalling (rhabdomeric- type photoreceptor cells)
- Cyclic nucleotide signalling (ciliary- type photoreceptor cells)

The phylogenetic relationship of the animal phototransduction cascade is shown in Figure 1.13.

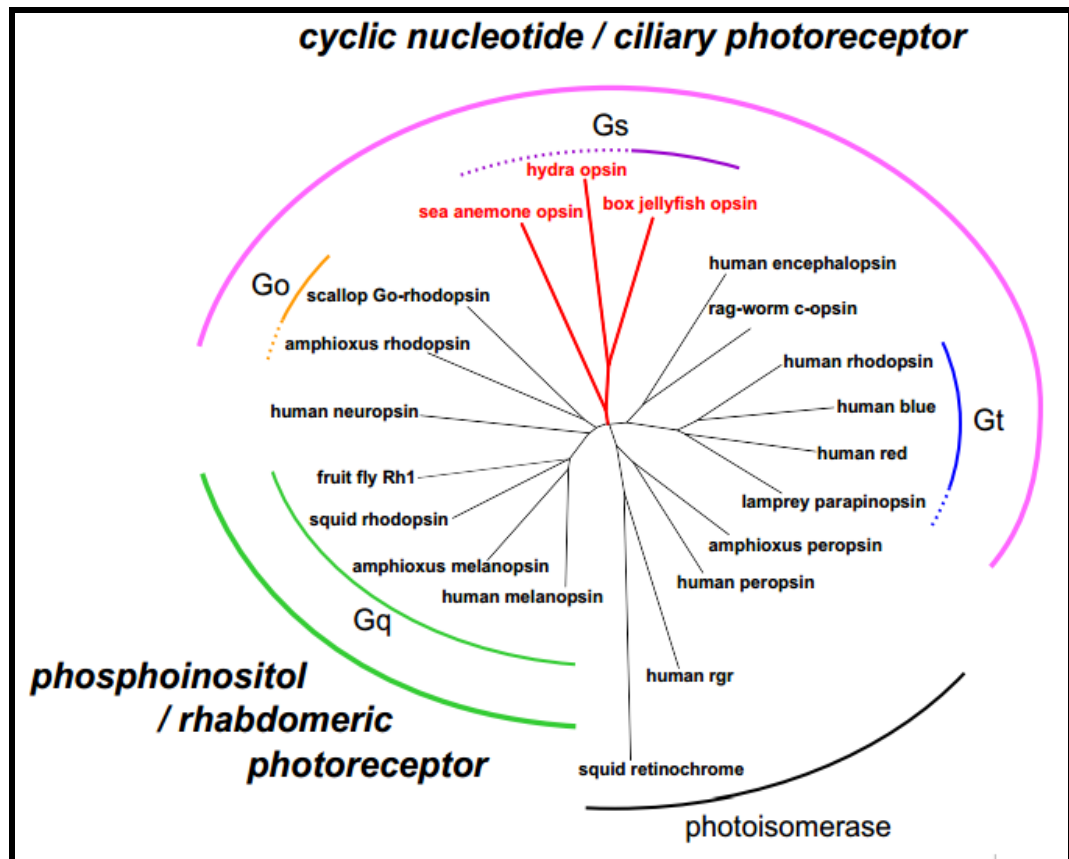


Figure 1.13 The animal's opsin phototransduction cascade (Koyanagi *et al.*, 2008)

The Figure above shows the phylogenetic relationship of opsins' phototransduction cascade among animals. The opsins that belong to Go, Gs and Gt subfamilies shown under the pink solid line use Cyclic nucleotide signalling (ciliary- type photoreceptor cells) pathway. The opsins that belong to the Gq subfamily use Phosphoinositol signalling (rhabdomeric- type photoreceptor cells).

1.10.3.1 Phosphoinositol Signalling (rhabdomeric type photoreceptor cells)

Phosphoinositol signalling of invertebrate visual pigments and melanopsin is carried out by Gq- coupled opsin based pigments inside rhabdomeric photoreceptors. Phosphoinositol signalling steps are shown in Figure 1.14 in the form of a flow chart (Koyanagi and Terakita, 2013)

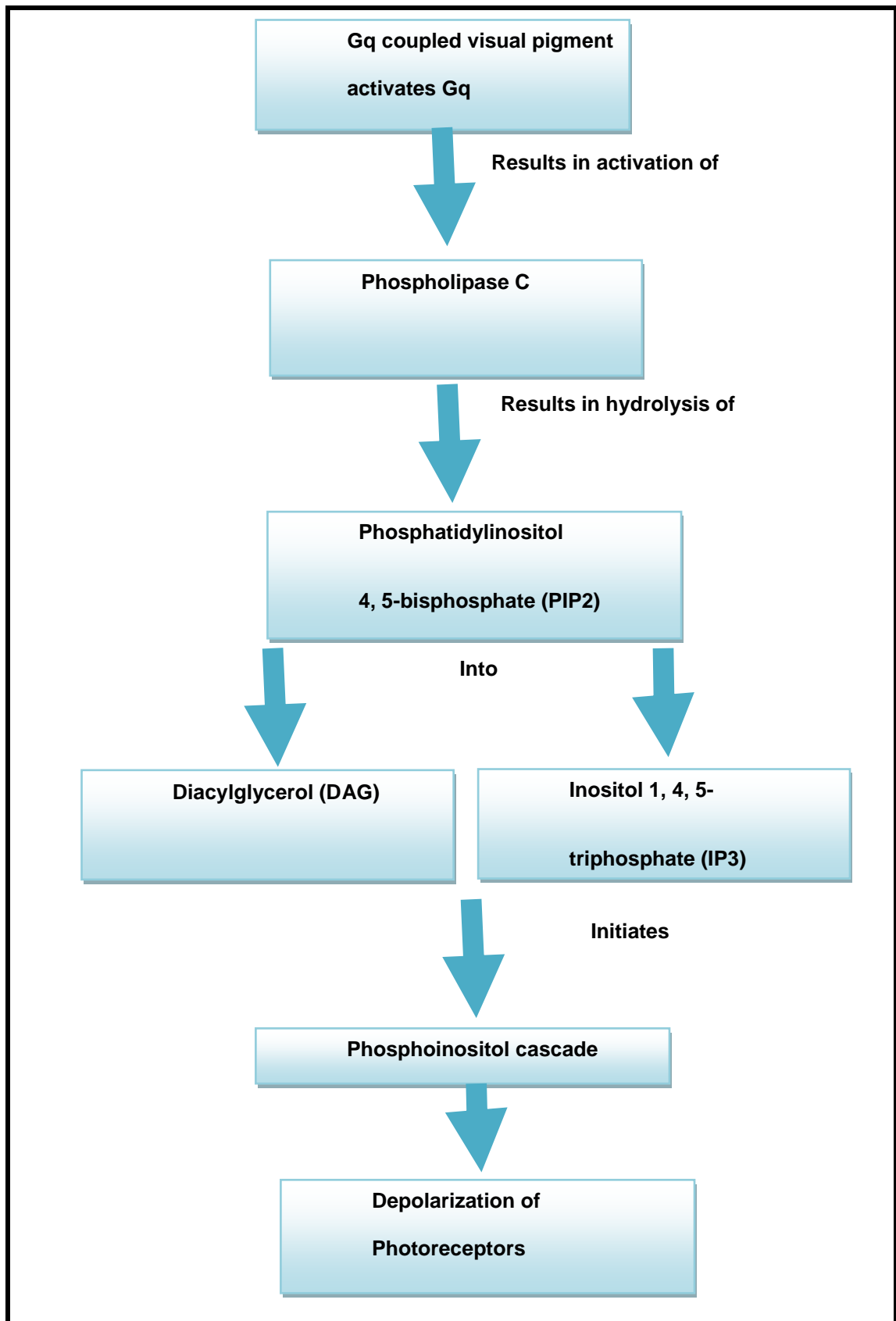


Figure 1.14 Steps involved in phosphoinositol signalling (Koyanagi and Terakita, 2013)

The above Figure shows various steps involved in phosphoinositol signalling pathway.

Melanopsin Opn4 is the vertebrate non-visual opsin, orthologous to invertebrate Gq-coupled visual pigments, activated by Gq type G protein phosphoinositol signalling cascade (Koyanagi and Terakita, 2013).

1.10.3.2 Cyclic Nucleotide Signalling (Cyclic type photoreceptor cells)

Cyclic nucleotide signalling involves a guanine nucleotide binding protein called “G protein” or “transducin”, which amplifies and transduces the light signal from opsin to inside of cell. The G protein is a heterotrimeric protein ($G\alpha\beta\gamma$) bound with GDP that is activated by exchange of guanine nucleotide from GDP to GTP and split into GTP bound $G\alpha$ and $G\beta\gamma$ subunits. The activated $G\alpha$ -GTP binds and activates the PDE (Phosphodiesterase). The activated PDE converts the cGMP into 5'GMP, by the breakage of a phosphodiester bond. This result in a decrease in the concentration of cGMP which in turn results in closing of CNG (cyclic nucleotide gated) channels and result in a hyperpolarization response in photoreceptor cells and release of neurotransmitters. The light signal finally reaches the brain through the optic nerves formed by ganglion cells (Shichida and Matsuyama, 2009). The mechanism of phosphotransduction and GPCR cyclic activation is shown in Figure 1.15a & b.

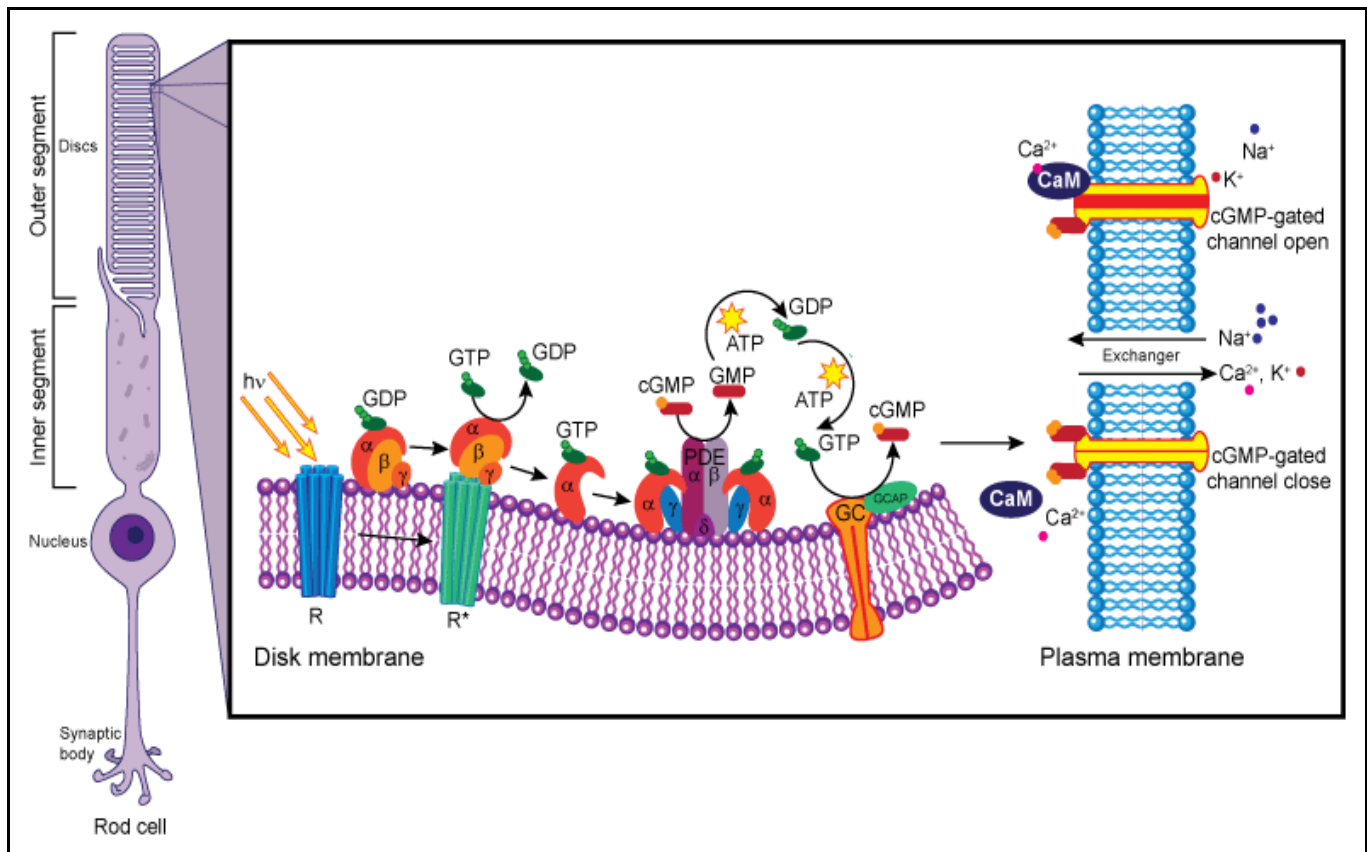


Figure 1.15 a The light and dark phototransduction cycle in the rod cell

(https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=282#rf)

The figure above illustrates that in the dark rhodopsin is in inactivated form and binds opsin with 11-cis-retinal in the disc membrane. The guanylyl cyclase (GC) keeps cGMP concentration high, and results in binding of Ca^{2+} with Ca^{2+} bound calmodulin and keeps cGMP-gated channel open. Both Na^{+} and Ca^{2+} enters the cell and results in high concentration of Ca^{2+} and Ca^{2+} -bound guanylate cyclase activating protein (GCAP). In light, the rhodopsin becomes activated by binding all trans-retinal, which in turn activates the heterotrimeric G ($\text{G}\alpha\beta\gamma$) protein to exchange GDP with GTP. Activated GTP bound $\text{G}\alpha$ separates from its $\text{G}\beta\gamma$ subunit and activates PDE (Phosphodiesterase). The activated PDE converts the cGMP into 5'GMP by the breakage of the phosphodiester bond. This results in a decrease in the concentration of the cGMP which in turn results in closing of cGMP-gated channel (Ridge *et al.*, 2003).

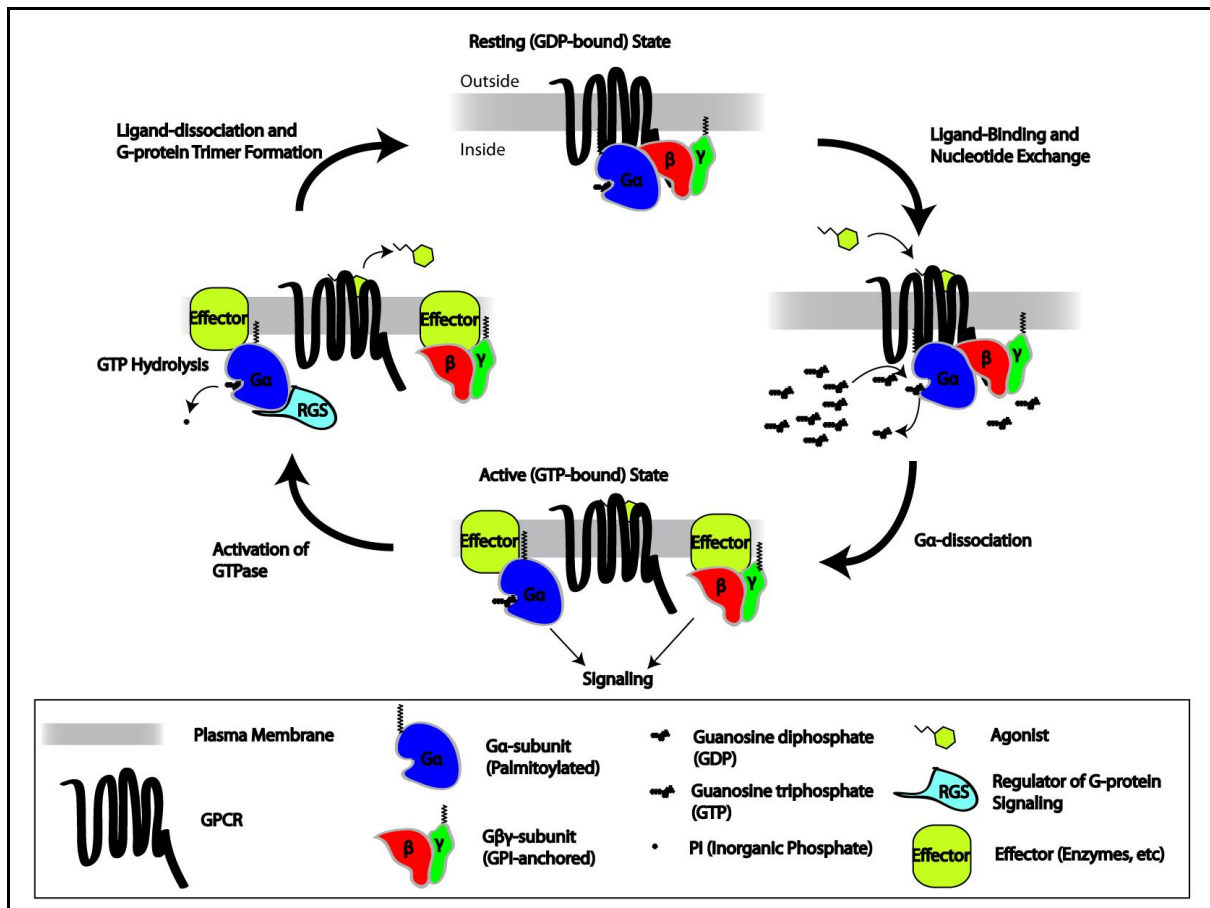


Figure 1.15 b The cyclic activation of GPCR G-protein

(http://upload.wikimedia.org/wikipedia/commons/c/c9/GPCR_cycle.jpg)

Figure shows the GPCR-G-protein cyclic activation/deactivation. In the resting state the GPCR receptor bound with a heterotrimeric G-protein when GDP bound to its alpha subunit. The photon of light induces conformational changes first in the GPCR receptor and then in the G-protein alpha subunit from G-beta/gamma subunits, results in release of GDP immediately replaced by GTP. GTP hydrolysis results in inactivation of the G-alpha subunit and once again may bind G-beta/gamma and form a heterotrimeric G-protein, which bound with inactive GPCR.

The classification of animal phototransduction cascade into groups and sub-groups on the basis of their functions is shown in Figure 1.16.

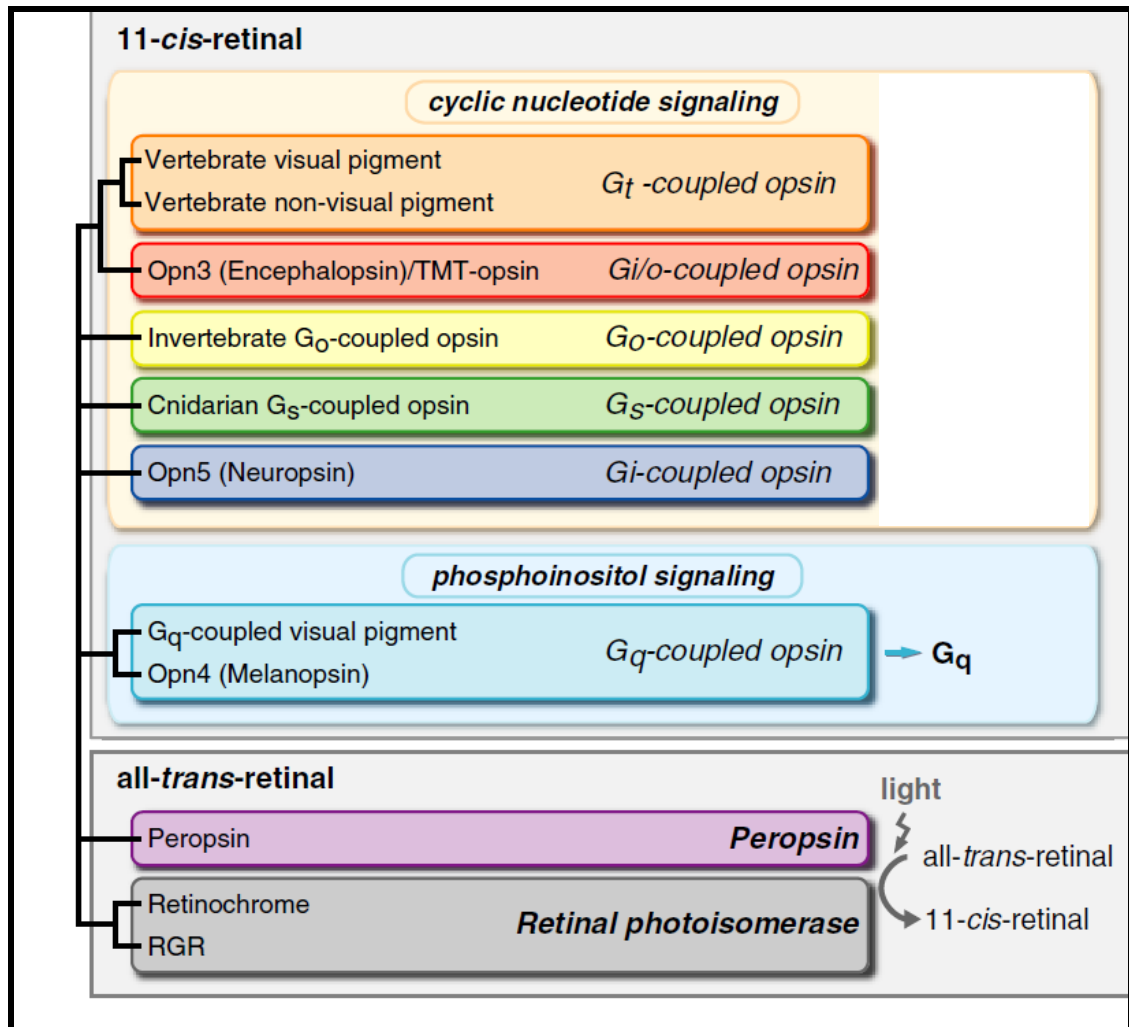


Figure 1.16 Signalling pathways functional diversity (Koyanagi and Terakita, 2013)

On the basis of functions, opsins are classified into 8 groups. The photopigment of six groups is formed by binding of opsin with chromophore 11-*cis* retinal that activates G protein mediated signal cascade. There are two main types of signalling cascades, 1) cyclic nucleotide signalling is triggered by *G_t*-coupled, *G_{i/o}*-coupled *G_o*-coupled, *G_s*-coupled and *G_i*-coupled opsin-based pigments and 2) Phosphoinositol signalling, triggered by *G_q*-coupled opsin-based pigments. Members of the last two groups, the peropsin and the retinal photoisomerase, bind with all-*trans*-retinal and light convert it into all-*cis*- form.

1.10.3.3 Types of Photopigments

The vertebrate photopigments are divided in two main categories: bleaching pigments and bistable pigments, discussed in Table 1.3 (Tsukamoto and Terakita, 2010; Koyanagi and Terakita, 2013).

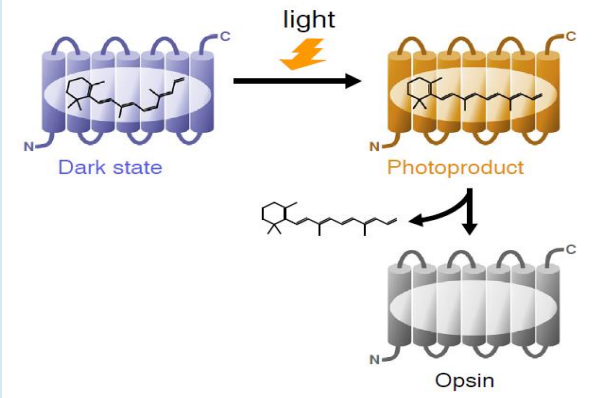
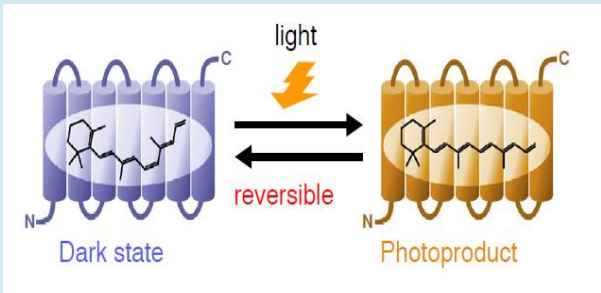
Bleaching Pigments	Bistable pigments
The activated form of vertebrate visual pigment photoproduct (metarhodopsin) is thermally unstable and releases its chromophore all-trans-retinal.	The activated form of photoproduct is thermally stable and converts back into original dark state by subsequent light absorption.
Example: Bovine rhodopsin	Example: Squid rhodopsin, parapinopsin
 <p>Figure explains the mechanism of bleaching pigment. The photoproduct is highly unstable and opsin releases its chromophore all-trans-retinal and become a colourless opsin.</p>	 <p>Figure explains the mechanism of bistable photopigments which can easily revert to the original dark state without releasing chromophore. Both states are thermally stable, the ground state as well as activated state.</p>

Table 1.3 Types of photopigments (Koyanagi and Terakita, 2013)

The table above explains the two types of photopigments: bleaching and bistable pigments.

1.10.3.3 The counterion

The counterion is an amino acid residue in rhodopsin that carries a negative charge, which balances the positive charge on the chromophore, retinaldehyde. The counterion, glutamic acid (E/ Glu) or aspartic acid (D/Asp), in vertebrate rhodopsin is present at position 113 in the third transmembrane helix. The same counterion Glu is present at position 181 in the second extracellular loops in invertebrate's rhodopsin, shown in Figure 1.17. The UV sensitive Parapinopsin has Glu at both positions (Terakita *et al.*, 2004).

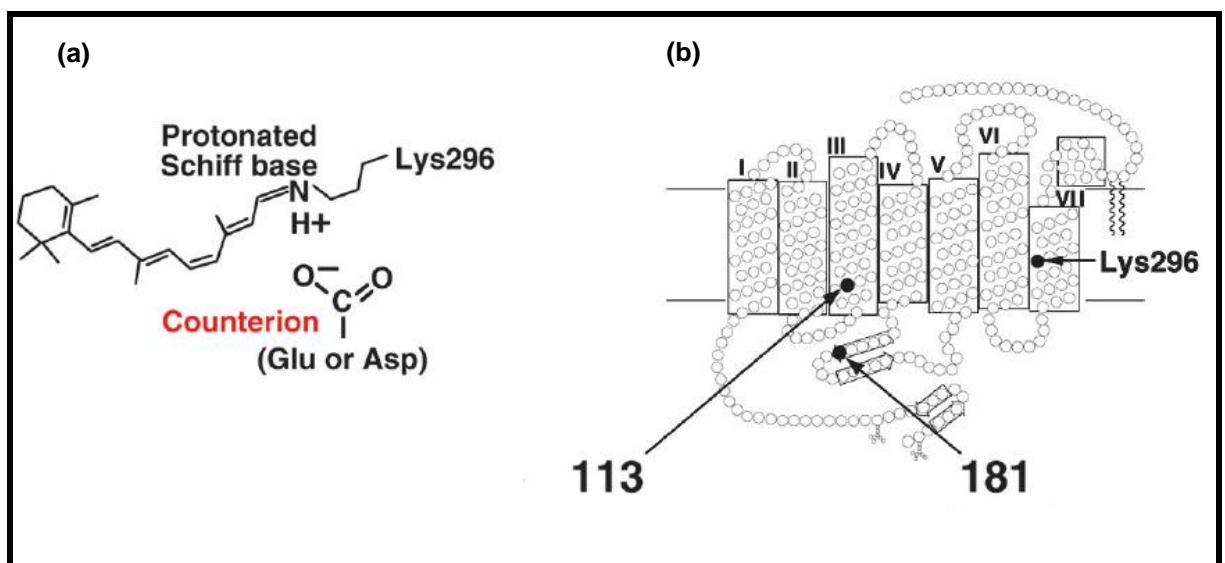


Figure 1.17 Binding of chromophore with Lys296 and position of counterions (Terakita *et al.*, 2004).

Part (a) of the above figure shows the binding of retinal chromophore with Lys296 through a protonated Schiff base in helix VII stabilized by the negatively charged counterion. Part (b) of the figure shows the position of counterions 113 and 181 in helix III and in the 2nd extracellular loop respectively, and Lys296 in helix VII.

1.11 Parapinopsin

Parapinopsin is UV sensitive non-visual bistable pigment expressed in pineal organs of lower vertebrates such as lamprey, trout, pike, frogs and green iguana (Koyanagi *et al.*, 2004; Seiji Wada *et al.*, 2012; Koyanagi and Terakita, 2013). Parapinopsin shows similarity with bovine rhodopsin, ~60%, (shown in Figure 1.18) although it has no great ability to activate G protein due to less relative conformational changes between helix VI to helix V as compared to rhodopsin in its active state (Tsukamoto *et al.*, 2009).

(a)		(b)				
Invertebrate Opsins		113	122	181	189	310 -312
rhabdomic		Y	F	E	F	HPK
melanopsin		Y	I	E	W	HPK
RGR		H	L	E	L	REM
cnidopsin		Y	I	E	P	NRS
encephalopsin		D	I	D	V	NRK
Chordate Ciliary Opsin						
parietopsin	Pt	Q	I	E	I	NKQ
parapinopsin	Pp	E	I	E	P	NKQ
VA	VA	E	I	S	P	NKQ
Ci-Opsin1	Ci	E	V	E	P	NRQ
LWS	LW	E	I	H	P	NRG
pinopsin	Pn	E	I	E	P	NKQ
SWS1	SW	E	L	E	P	NKQ
SWS2	B	E	M	E	P	NKQ
Rh2	G	E	Q	E	P	NKQ
RhA lamprey		E	E	E	P	NKQ
Rh1 jawed	Rh	E	E	E	I/V	NKQ

Figure 1.19 Important amino acid positions in Parapinopsin compared with other opsins (Lamb, 2009)

The Figure shows the type of amino acid in rhabdomic opsins and ciliary opsins. The amino acid residue at position 113 and 181 is called a counterion. In part "a" of the figure the names of all opsins are given while in part "b" of the figure the table shows the position and names of important amino acids residues. The invertebrate opsins rhabdomic, melanopsin, RGR, cnidopsin and encephalopsin have counterion at position 181 whereas Parapinopsin, VA and pinopsin are vertebrate non-visual opsins and have counterions at positions 113 and 181 except parietopsin which has a counterion at position 181 like invertebrate opsins. The rest of the chordate ciliary opsins are visual opsins.

1.11.2 Evolution in Position of Counterions and Parapinopsin

Phylogenetic studies show that primitive opsins had a counterion for a Schiff base at position 181 (E181) as in invertebrate opsins. Evolutionary studies show that ciliary chordate opsins have evolved a second negatively charged residue at position 113 (E113) shown in Figure 1.18, which stabilizes the ground state/resting state configuration while E181 helps in the stabilization of active/metarhodopsin state, when opsin binds with all trans retinal (Lamb, 2009).

Parapinopsin is a key pigment in understanding the evolution of vertebrate visual pigments. The vertebrate visual pigment has counterion/glutamic acid at position E113 while invertebrate visual pigment has the counterion glutamic acid at position E118, and Parapinopsin has counterions at both E113 and E118 positions. The ability to activate G protein is much less for Parapinopsin, like invertebrate visual pigments. These facts suggest that vertebrate's visual pigments evolved from ancestral vertebrate non-visual pigments like Parapinopsin (Yamashita *et al.*, 2011; Tsukamoto and Terakita, 2010).

1.11.3 Visual-Arrestin and β -Arrestin

Arrestin is the key protein which terminates the G protein mediating signalling. There are two types of arrestin in lower vertebrates and mammals, discussed in Table 1.4.

Visual arrestin	β -arrestin
Visual arrestin binds with light stimulated visual pigments and shuts off the G protein signalling resulting in phototransduction inactivation	β -arrestin binds with non-visual GPCR <i>i.e.</i> Parapinopsin in pineal photoreceptor cells. The β -arrestin plays an additional role other than termination of G protein activated cascade. That is the removal of receptors from the cell surface through calthrin mediated internalization due to the presence of a calthrin mediated domain, which is not present in visual arrestin proteins.

Table1.4 Types and functions of arrestin (Yamashita *et al.*, 2011)

First column in the Table explains the function performed by visual arrestin, whereas the 2nd column explains the function performed by β -arrestin.

1.11.4 Monostable Vs Bistable Pigments G Protein Activation

Light induces conformational changes in rhodopsin to activate G proteins. This G protein activation efficiency is (1/20–1/50) times lower in bistable pigments *e.g.* vertebrate non-visual pigments (Parapinopsin) or invertebrate visual pigments. In bovine rhodopsin the transmembrane cytoplasmic end of helix number VI is important for G protein activation. Helix VI moves towards helix V and away from helix III. The “helix VI movement” is the characteristic feature whether they belong to monostable or bistable pigments. Site directed mutagenesis studies reveal that “helix VI movement” in Parapinopsin is smaller when compared to bovine rhodopsin, shown in Figure 1.20. Hence the greater the angle of helix VI movement, the higher will be the G protein activation efficiency (Patel *et al.*, 2004; Tsukamoto and Terakita 2010; Tsukamoto *et al.*, 2010; Shichida and Yamashita, 2003)

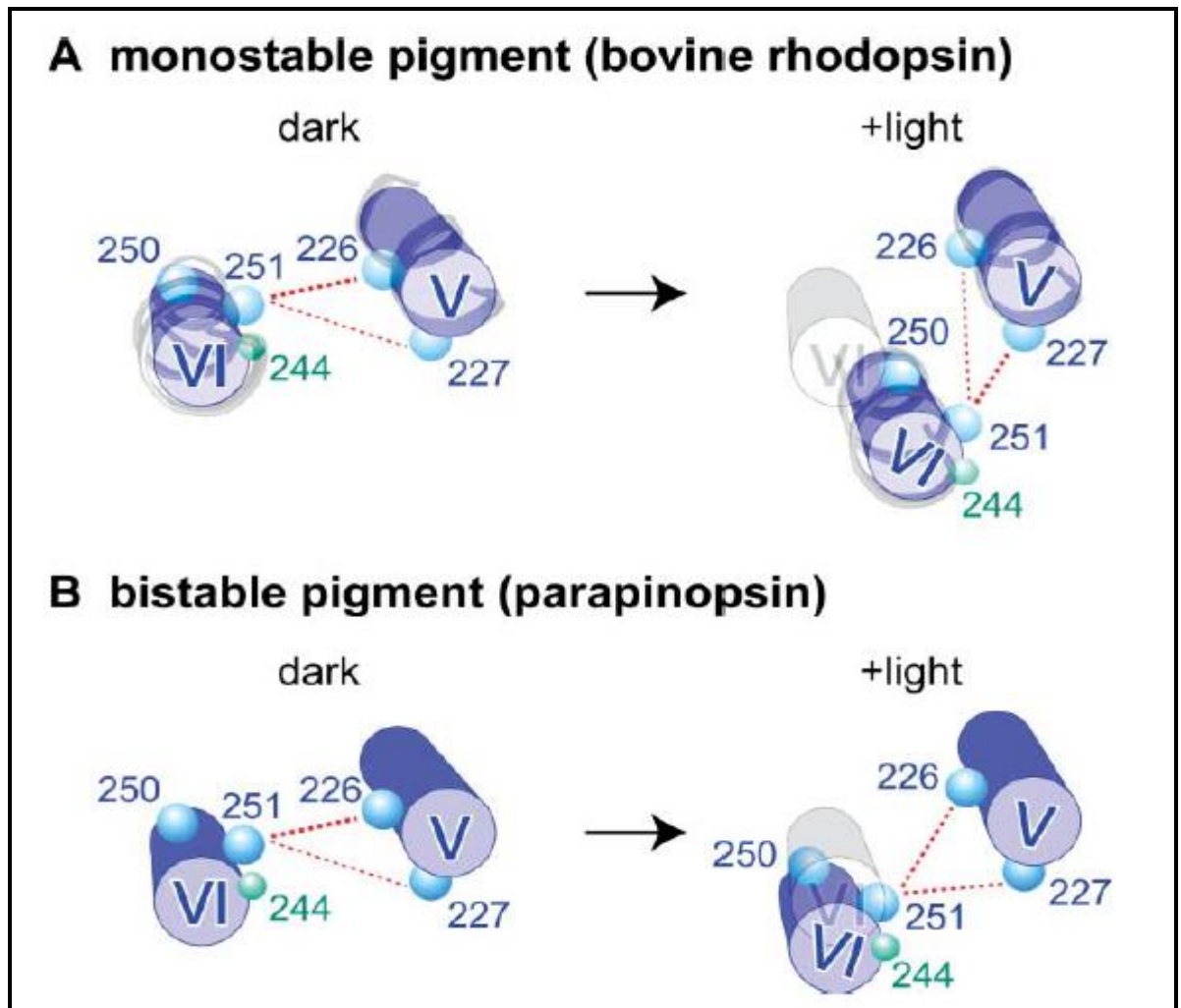


Figure 1.20 Conformational changes in helix VI and V in bovine rhodopsin and Parapinopsin (Tsukamoto and Terakita 2010)

(a) and (b) of the figure shows the model for bovine rhodopsin and Parapinopsin respectively. Helix VI movement as compared to helix V in the dark state and active state is shown. In the active form the presumed state of helix VI is shown in grey, while the shift in helix VI is shown in blue. The shift in position of helix VI is greater in bovine rhodopsin as compared to Parapinopsin.

1.12 Work Done on Parapinopsin

1.12.1 Catfish Parapinopsin

In 1997, Blackshaw and Snyder discovered for the first time a member of a new gene family: a novel opsin named “Parapinopsin” and expressed in the parapineal organ of catfish.

1.12.2 Lamprey Parapinopsin

RT-PCR studies have shown the expression of a UV sensitive Parapinopsin in lamprey pineal along with rhodopsin (Yamashita *et al.*, 2007).

1.12.2.1 Types of Ganglion cells in Pineal organ

Two types of ganglion cells are present in lower vertebrates’ (lamprey, teleost and frogs) pineal organ, and show antagonistic chromatic and achromatic responses.

Chromatic ganglion cells	Achromatic ganglion cells
Neural activity of chromatic type of ganglion cells is inhibited by short wavelength/UV light called antagonistic response and excited by long wavelength/green light (540nm).	Neural activity of achromatic type of ganglion cells is inhibited by long wavelength/ green light and excited by shortwave length (525nm).

Hence pineal organ can detect wavelength discrimination due to the presence of UV sensitive and green light sensitive photoreceptor cells (Yamashita *et al.*, 2007).

1.12.3 β Arrestin and Lamprey Pineal

Yamashita and his group in 2011 found that β arrestin bound Parapinopsin in cultured cells, opposite to visual opsins that bind visual arrestin.

1.12.4 Pineal the UV Reception Organ in Vertebrates

Koyanagi *et al.*, in 2004 isolated UV sensitive Parapinopsin homologues from fish (lamprey, rainbow trout) and frog (*X. tropicalis*) pineal complexes.

1.12.5 Expression of Parapinopsin in Reptile Lineage

Wada *et al.*, in 2012 found for the first time the expression of Parapinopsin in the parietal eye of green iguana. Their Western Blotting results showed the 38kDa Parapinopsin protein expressed in the parietal eye. Parapinopsin wavelength discrimination is present in reptiles in a pineal related organ, parietal eye. Parietopsin is a non-visual opsin having visual and non-visual opsin characteristics, expressed in the parietal eye (Sakai *et al.*, 2012). S Wada's group found the green anole Parapinopsin gene sequence in the genome database to design primers, and used these primers against green iguana and successfully got iguana Parapinopsin (Wada *et al.*, 2012).

1.12.6 Position of Parapinopsin in Phylogenetic Tree

Figure 1.21 shows a phylogenetic tree and represents the position of Parapinopsin among different opsins. The method used by Wada *et al.*, 2012 to construct the tree was "Neighbour Joining Method" based on evolutionary distance data used by Saitou and Nei in 1987. The protein sequences of all opsins was aligned by MAFFT (Multiple Alignment Fast Fourier Transform) to determine the evolutionary history of the protein family on the basis of the presence of conserved amino acid sequences (Kato *et al.*, 2002). MAFFT is a high speed multiple sequence alignment program available at EMBI/EBI.

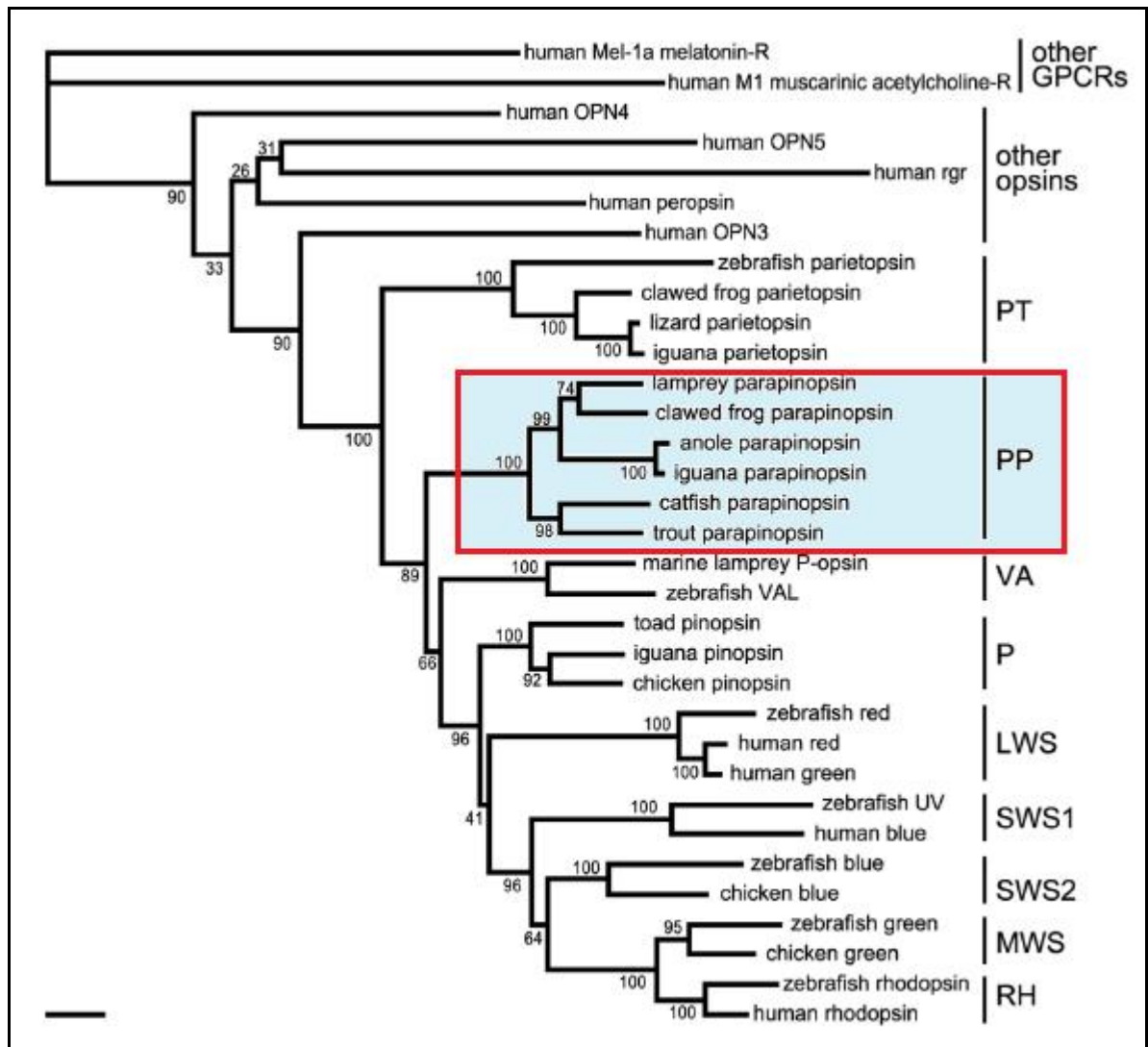


Figure 1.21 Position of Parapinopsin among other opsins in a phylogenetic tree (Wada *et al.*, 2013)

The figure shows the phylogenetic tree constructed on the conserved amino acid alignment basis. Each branch node has bootstrap values (bootstrap is a computer based statistical assessment which determine the accuracy of data *i.e.* higher bootstrap value indicates strong support for a clade Efron *et al.*, 1996). The name and group of each opsin is given at the right side of each cluster, and the scale bar represents the 0.1 substitution per base. "PP" represents the Parapinopsin position in the phylogenetic tree, highlighted blue inside the red box. All the sequences were taken from DDJB/EMBL/Genbank databases.

1.13 Amphibian Species Used during Current Studies & their

Genetics

To study Parapinopsin the amphibian species of *Xenopus tropicalis*, *Xenopus laevis*, *Rana temporaria* and *Rana tigrina* were used. The information about frog species is given in Table 1.5.

Features	Species			
	<i>X. tropicalis</i>	<i>X. laevis</i>	<i>R. temporaria</i>	<i>R.tigrina</i>
Common name	Western clawed frog	African clawed frog	Common European frog	Asian bullfrog
Ploidy	diploid	allotetraploid	diploid	diploid
Haploid chromosome number	10	18	13	26
Genome size (bp)	1.7×10^9	3.1×10^9	4.3×10^9	4.4×10^9
Adult size (cm)	4-5	10	6-9	12-18

Table 1.5 Information about Frog species used during current studies

The table provides different characteristic information (column1) about *X. tropicalis*, *X. laevis*, *R. temporaria* and *R.tigrina* (column 2,3,4 and 5 respectively). (Hirsch *et al.*, 2002; Cano *et al.*, 2011; MacCulloch *et al.*, 1996; Kotpal, 2010; Saba and Tripathi 2013; <http://www.genomesize.com/results.php?page=1>)

1.13.1 Classification of Frogs

Table 1.6 gives the details about *X. tropicalis*, *X. laevis*, *R. temporaria* and *R.tigrina* classification.

Units of Classification	<i>X. tropicalis</i>	<i>X. laevis</i>	<i>R. temporaria</i>	<i>R.tigrina</i>
Kingdom	Animalia	Animalia	Animalia	Animalia
Phylum	Chordata	Chordata	Chordata	Chordata
Subphylum	Vertebrata	Vertebrata	Vertebrata	Vertebrata
Class	Amphibia	Amphibia	Amphibia	Amphibia
Order	Anura	Anura	Anura	Anura
Family	Pipidae	Pipidae	Ranidae	Ranidae
Genus	<i>Xenopus</i>	<i>Xenopus</i>	<i>Rana</i>	<i>Rana</i>
Species	<i>Xenopus tropicalis</i>	<i>Xenopus laevis</i>	<i>Rana temporaria</i>	<i>Rana tigrina</i>

Table 1.6 Classification of *X. tropicalis*, *X. laevis*, *R. temporaria* and *R.tigrina*

The Table explains the classification of all frogs used during current study. *X. tropicalis* and *X. laevis* belongs to the family Pipidae (Tongueless frogs), while *R. temporaria* and *R.tigrina* belongs to the family Ranidae (true frogs). Haque and Saidapur 1994, Daudin, 1802; <http://animaldiversity.ummz.umich.edu/> and <http://clawedfrogs.tripod.com>)

1.14 Aim of Project

Parapinopsin is a novel, non-visual; UV sensitive protein which belongs to the “Opsin” family of G-protein coupled receptors (GPCR). Parapinopsin is expressed in lower vertebrates and reptiles, in the pineal and pineal related organs. GPCR act as receptors for various ligands like hormones, neurotransmitters, ions and photons. GPCR link the cell’s environment with the external environment. GPCR make the largest family of proteins with nearly 800 members known in human-beings. Latest research reveals the importance of opsin-based pigments due to the “light switch” properties of these molecules, which can be used for target cells to induce particular physiological responses by exogenous stimuli for example the light instead of chemicals (Koyanagi and Terakita, 2013).

The aim of the current project was to study amphibian Parapinopsin and to learn more about Parapinopsin as very little information is available about this protein in amphibians. The first part of the current project was to study amphibian Parapinopsin genetics. Four species of frogs were selected by using local resources and by contacting different groups who are already working on different frog species. Two frog species out of four belong to the *Pipidae* and the rest belong to the *Ranidae* family. Dr Bell Esther group (King’s College London NHH) is working on *Xenopus laevis* and made this animal available. *Xenopus tropicalis* cDNA/gDNA was commercially purchased from Portsmouth. Access to *Rana temporaria* and *Rana tigrina* was made possible by the help of Dr Peter Eagles (London, UK) and Mrs Iffat Rizvi (Pakistan), respectively.

The second part of the project was aimed at studying Parapinopsin protein from different frog’s species. To accomplish this; Parapinopsin specific antibodies were generated against peptides for Western Blot analysis. The Western Blot study indicates the presence and the Parapinopsin protein size. To get further information about Parapinopsin protein, expression studies was conducted. The codon optimized Parapinopsin gene was commercially synthesized and cloned in GPCR specific plasmids called pASK-GPCR plasmids (kind gift from Professor Georgiou, USA).

The third part of project was based on a tissue culture based “melanophore” study. *X. laevis* melanophore cell lines were cultured to extract the RNA. The expression of Parapinopsin mRNA in melanophores was studied by Northern Blotting.

The fourth part of the project was a Bioinformatics based study. Different Bioinformatics tools/software were used to predict the 2D, 3D Parapinopsin protein structure and to generate phylogenetic tree which will represent the position of Parapinopsin among other types of opsins.

CHAPTER 2 MATERIAL & METHODS

This chapter contains accounts of different molecular biology methods employed in this thesis along with the materials required for the work.

2.1 RNA Extraction and DNA Purification

Some spare embryos of *Xenopus laevis*, *Rana temporaria*, and *Rana tigrina* were kindly provided by Jacqueline Tabler, Guy's Tower, Dr. Bell Esther, New Hunt's House King's College London, (Dr. Peter Eagle's) garden water pond and Pakistan PMAS-UAAR respectively. *Xenopus tropicalis* gDNA/cDNA was commercially purchased from University of Portsmouth. Embryos were kept and grown to stage 53 (appendix 3) and used to assay for mRNA expressed in the pineal. RNA extraction was carried out by using the RNeasy Mini Kit (Qiagen, UK), as described in the manufacturer's protocol. To remove traces of DNA contamination, extracted RNA was further treated with DNase by using Ambion® TURBO DNA-free™ (Ambion Applied Biosystems, UK), according to the manufacturer's instruction. Genomic DNA from tadpoles was extracted by using QIAamp DNA Mini Kit (Qiagen, UK), according to the manufacturer's instructions.

2.2 Quantification of DNA and RNA Concentration

To measure DNA/RNA concentrations, Optical density (OD) at wavelength 260nm was measured by spectrophotometry, GeneQuant (Pharmacia Biotech, Sweden). First the spectrophotometer was adjusted at 260 nm by using a clean quartz cuvette (1 cm light path) with 100µl of sterile water to give a zero reading. Then a 1µl of DNA/RNA solution was diluted in 99µl of water and the OD reading recorded. The concentration of DNA/RNA was calculated by use of the following formula (Table 2.1)

RNA	DNA
1.0 OD ₂₆₀ = 44ng/ µl	1.0 OD ₂₆₀ = 50ng/ µl

Table 2.1 Formulae used to measure RNA and DNA concentrations

OD260/ OD280 ratio was used to determine the purity of samples. For highly pure DNA and RNA samples these ratios are 1.8 and 2.0 respectively.

2.3 Synthesis of cDNA by RT

Complementary DNA (cDNA) is the DNA synthesized from an existing RNA strand when it is used as a template by reverse transcriptase; a process called reverse transcription (RT). In eukaryotes, owing to 3' polyadenylation of messenger RNA, a polydeoxythymidine oligomer of 16-18 bases can be used as primer to produce RNA-DNA hybrids. To obtain single stranded DNA from this DNA-RNA hybrid, an endoribonuclease enzyme RNase H is used that only cleaves RNA/DNA hybrids, allowing amplification by the polymerase chain reaction (PCR). 45µl of total RT reaction mixture contained: a volume equivalent to 1-2µg of RNA, 1µl of oligo dT16-18 (2µg/µl, Synthesized by the Molecular Biology Unit in the Hodgkin Building, King's College London), and made up to the total volume using RNase free water (Qiagen, UK). To denature the secondary structure of RNA, the mixture incubation was carried out at 70°C for 10 minutes in (Thermal Reactor- HYBAID Omnigene, UK), followed by a 5 minute incubation on ice. The following were added to get 100µl final volume of RT mixture: 5µl of 10mM deoxynucleotide triphosphate (dNTP) (Promega, UK), 20µl of first strand 5x Buffer (250mM Tris-HCl: pH 8.3, 375mM KCl, 15mM MgCl₂: Invitrogen, UK), 10µl of 1M DTT (Invitrogen, UK), 5µl of RNasin Ribonuclease Inhibitor, 20-40U/µl (Promega, UK), and 5µl of M-MLV reverse transcriptase, 200u/µl (Promega, UK). Afterwards the mixture was incubated at 37°C for one hour (Techne DB, DRI-Block®, U.S.A).

2.4 Primer Designing

The cDNA sequence of interest was retrieved from NCBI catalogued under EST (Expressed Sequence Tag) database used to design cDNA specific PCR primers. The length of primers was between 20-22 bases with 50% GC. BLAST search for similarities of Primer sequences was carried out, by comparing selected primer sequences to the sequences present in the database by using Primer-BLAST at NCBI. Primers were

made by IDT (Integrated DNA technologies, UK). The primer sets used in the project are shown in Table 2.2.

PCR Primer	Sequence of primer set	Product size (bp)	Annealing Temperature
PNPexon2 (I)	Forward 5' d-aatgggcacactcacattcacc3' Reverse 5'd- gatgtaggacatgttgactggg 3'	181	60°C
PNPexon2 (O)	Forward 5'd-gcaccattgcagtaatagcgtt3' Reverse 5'd- gaggaatgggatggcaaagca 3'	269	61.5°C
PNP1	Forward 5'd -aatgggcacactcacattca3' Reverse 5'd- cttgcacctcggtttactc3'	320	60°C
Fcdna	Forward5'd -tgttgccctgacaatttcggtg3' Reverse 5'd-gccggttatactggagataaaag3'	1117	60°C
PNPd	Forward5'd-ctgtgccgccgctctctttctc3' Reverse5'd-gtaggcgttccactcgttacgctgata3'	900	59.1°C
Actin xl	Forward 5'd -catacagccatcttgcttac3' Reverse 5'd-gttgtgggagattggcaag3'	231	60°C
β-actin	Forward 5'd -ggacctctatgccaaactgttc3' Reverse 5'd -tcagcaattcctgggtacatt3'	60	60°C

T7	5'd-taatacgactcactataggg3'	50°C
SP6	5'd-atttaggtgacactatag3'	50°C

Table2.2 Primer sets used within the project, 5'→3'

The table shows the abbreviation of primers in the 1st column, sequence information in the 2nd column, product in (bp) in the 3rd column and annealing temperature in the 4th column. *X. tropicalis* Parapinopsin cDNA sequence in appendix 7 indicated the regions where primers prime.

2.5 Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a high temperature amplification of a DNA segment directed by a sequence specific set of primers. The initial denaturation of the DNA template is at 95°C; primer annealing temperature to the template is at 55-60°C and extension of primers by a heat stable DNA polymerase takes place at 72°C. A thermocycler repeats this temperature cycle many times according to a pre-set program. The first amplified product is used as a template for the next amplification and results in an exponential increase in the amount of product. Hot start PCR master mix (Promega, UK) was used for all PCR reactions.

Reaction Mixture components	Volume Per tube (25µl)
20ng of DNA	xµl
Forward Primer (10µM)	1µl
Reverse Primer (10µM)	1µl
PCR Master Mix	12.5µl
Nuclease free H2O	Add to 25µl

Table2.3 (a) Recipe for PCR Master Mix (25µl reaction mixture)

The PCR cocktail (Table 2.3 a) was prepared on ice. The cocktail and DNA, to final volume of 25 µl, were added to each PCR tube, and tubes were given a short spin before placing in the Thermocycler. The general profile, programmed by the thermal cycler, is shown Table 2.3 b:

1 cycle of Initial template denaturation at 95°C for 5 minutes			
	Actual step	Temperature	Time
30 cycles	Denaturation	95°C	1min
	Annealing	Depend upon the primer set used. (See in table 2.2)	30 sec
	Elongation	72°C	Depend upon the size of amplicon used, x (1 Kb/min)
Final extension at 72°C for 10 min			

Table 2.3 (b) Thermo cycler's PCR program for 30 cycles

2.6 Sequence Specific Primer- Polymerase Chain Reaction (SSP-PCR)

The SSP-PCR technique is sequence specific primer- polymerase chain reaction amplification and is a type of PCR reaction that relies on one or two 3' mismatches. This 3'-mismatch principle is used for the detection of a single nucleotide mutation e.g. SNP (single nucleotide polymorphism). For SSP-PCR, one needs to design two forward primers and one common reverse primer. One gene specific forward primer will give amplification and there is no amplification with the 2nd gene specific forward primer with the single 3' mismatched nucleotide. (Welsh and Bunce, 1999). The figure 2.1 below explains the SSP-PCR technique in detail. The SSP-PCR primer sequences are given in Table 2.4.

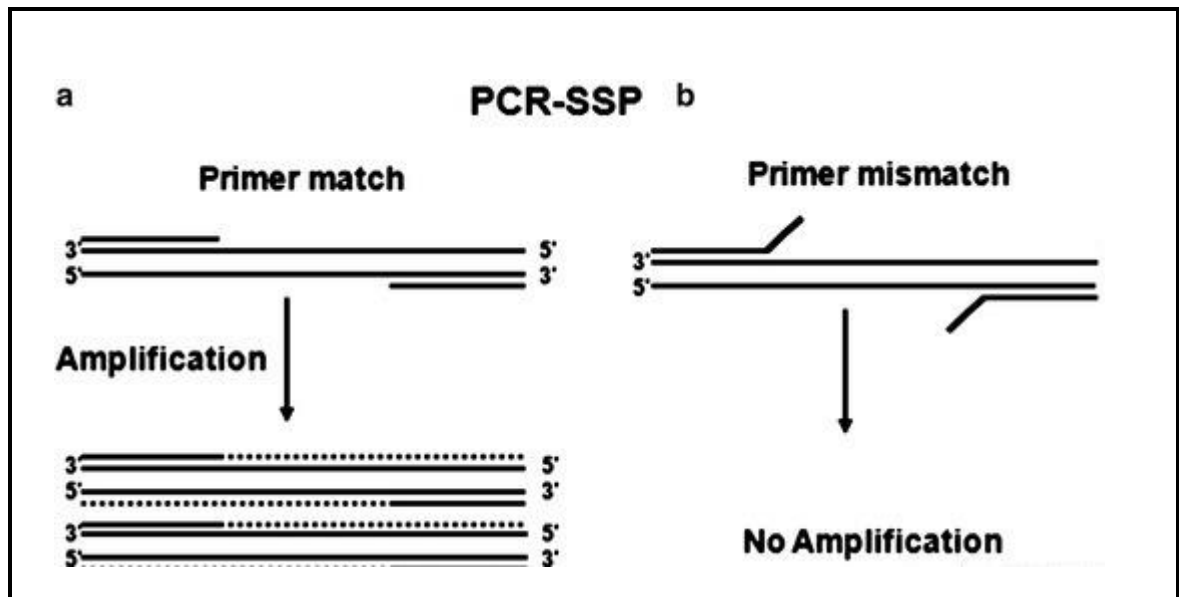


Figure 2.1 SSP-PCR technique

The PCR reaction was carried out by using two sets of primers. In part "a" The gene specific forward primer will give an amplified product. In part "b" the primer with 3' mismatched nucleotide gives no amplification.

(<http://www.dxy.cn/bbs/image/preview?uid=28339786&path=2013/04/07/01&topicId=25314936>)

2.6.1 SSP-PCR Primer Sequence

Four SSP-PCR forward primers and one common reverse primer were designed to study the point mutation at the 2nd position of the 156th codon which is G to A in Parapinopsin cDNA of *Rana temporaria* and *Rana tigrina*. The nucleotides in red (Table 2.4) are mismatched nucleotides at the 3'end. Primers were made by IDT (Integrated DNA technologies, UK).

SSP-PCR Primer	Sequence of primer set	Annealing Temperature	Product size (bp)
SSP1-pnpG	Forward5'dcggcattgctgcttctggatctg3'	63°C	181
SSP1-pnpA	Forward5'dcggcattgctgcttctggatcta3'	62.5°C	181
SSP2pnpTG	Forward5'dcattgctgcttctggatttg3'	62°C	181
SSP2pnpTA	Forward5'dcattgctgcttctggattta3'	61.5°C	181
Common reverse primer pnpR	Reverse5'dtgctttgccatcccattctc 3'	60 °C	

Table 2.4 SSP-PCR Primer sets used within the project, 5'→3'

The Table shows the abbreviation of the primer in the 1st column, its sequence in the 2nd column, the annealing temperature used for PCR in the 3rd column and the expected product size (bp) is shown in the column 4.

2.7 Visualization of PCR products

Agarose gel electrophoresis was used to analyze PCR products. A negative charge is present on the DNA molecule due to its phosphate backbone. Therefore during electrophoresis DNA molecules will migrate towards the positive pole. A network of pores is present in an agarose gel through which DNA molecules travel. The gel pore size is determined by the concentration of agarose, and decreases with the increase in agarose concentration. 1% and 2% agarose gels were prepared by dissolving 0.5g and 1g of agarose (Fisher Bioreagents, UK) respectively in 50ml of 1x TAE buffer (40mM Tris-Acetate-EDTA, 1mM EDTA pH 8.5, Sigma, Germany). 10µl of ethidium bromide (1mg/ml, MBL International, USA) was added and mixed, after melting the mixture at a maximum power for 2 minutes in a microwave oven. The gel apparatus was adjusted by placing a boat with comb horizontally in the electrophoretic tank (Max fill, Fisher brand, UK) and pouring the gel. Once the gel was set, the boat direction was changed to an angle of 90°.

The comb was removed and the tank was filled with 1x TAE buffer up to a mark in the tank. 6x loading buffer was mixed with the DNA samples and was loaded into the wells. (Loading buffer contained 0.25% bromophenol blue and 30% glycerol in 1x TAE buffer, BioLabs Inc, England). Suitable molecular weight markers (300-450ng) were used for size estimation and were also loaded onto the gel. The tank was supplied with a voltage between 80-90 volts for electrophoresis by connecting to a power supply (Pharmacia, UK). Bromophenol blue dye in the loading buffer runs at the fastest rate due to its low molecular weight therefore it runs in front of the sample. When the bromophenol blue dye reached 2/3rd of the gel length, the power supply was turned off and the gel observed under soft UV light (Gene flash, Syngene Bio Imaging, USA), and an image was taken (Video graphic printer, UP-850, Sony, 70 Japan).

2.8 Extraction of PCR Products

The required size of DNA fragments were cut out from an agarose gel. Fragments were then purified by using Wizard® SV Gel and PCR Clean-Up System. The protocol was followed according to the manufacturer's instructions (Promega, UK). Afterwards, the concentration of purified DNA fragments was measured by means of an OD260nm measurement.

2.9 DNA Ligation

Taq polymerase adds 'A' overhangs at the 3' ends of PCR products. Purified PCR products were directly ligated into the pGEM®-T Easy Vector (Promega USA), a linearized vector with a single 3'-terminal thymidine at both ends. The efficiency of PCR product ligation greatly increases due to the presence of T-overhangs at the insertion point and prevents the recirculization of the vector.

Following gel purification of PCR products, the concentration of DNA was estimated by means of OD260nm. For ligation, a molar ratio of insert to vector 10:1 was used. For this, the equation used was: $\text{ng of insert} = (\text{ng of vector} \times \text{kb of insert size} / \text{kb of vector size}) \times (\text{molar ratio of insert: vector})$. The total volume of the ligation reaction was 10 µl, with 200ng of the vector, xng of the insert, 1 µl of T4 DNA ligase enzyme (3U/ µl), 1 µl of 10x ligase buffer (Promega, UK) and a volume of xµl of DEPC treated water was used. The

ligation reaction was left at 4°C overnight. To confirm ligation, a PCR reaction was carried out by diluting the ligation reaction mixture 1000 times, which was used as template. Gene specific and plasmid specific primers were then used either both located within the plasmid or one within the plasmid and the other within the insert.

2.10 Preparation *E. coli* Competent Cells

Competent cells were made by using the *E. coli* strain XL1Blue (kindly provided by Molecular Biology Unit, Randall Centre, King's College London).

Frozen *E. coli* cells were thawed on ice, and inoculation of the original culture was done in 10ml of LB broth (10g of LB Broth, Sigma, UK dissolved in 500ml of deionized water and autoclaved) media. Cells were grown overnight at 37°C with continuous shaking. Fresh media was prepared for dilution of overnight grown culture by 1:100, and *E. coli* were grown till OD₆₅₀= 0.2-0.4. The culture was incubated on ice for 10 minutes and cells were spun at 5000 rpm at 4°C for 10 minutes and cells resuspended in 50ml of 100 mM MgCl₂ followed by 5min incubation on ice. These cells were then centrifuged at 4000rpm for 10 min at 4°C, and the pellet of cells was resuspended in 50ml of 100 mM CaCl₂; followed by 20 minutes incubation on ice. After that, the cells were centrifuged at 4000rpm/ 10 minutes at 4°C. The cell pellet was resuspended in 4ml of 100 mM CaCl₂ and 15% glycerol. Aliquots were made on ice and stored at -80°C.

2.11 Transformation

A 50 µl aliquot of XL1Blue competent cells was removed from the -80°C freezer and thawed on ice for 5 minutes. 5 µl of the ligation reaction mixture was added in an Eppendorf tube containing 50 µl of competent cells on ice, and left on ice for another 20 minutes. The competent cells were heat shocked on a heat block at 42 °C (Grant BT1 Block Thermostat, UK) for 50 seconds then put on ice for 2 minutes. To the cells, 350 µl of SOC medium (BioLine Ltd, UK) was added, followed by incubation at 37 °C for 45-90 minutes. 100 µl of transformed competent cells were spread on prepared LB agar plates. LB agar plates were prepared by using 16g of LB agar (Sigma, UK), dissolved in autoclaved deionized 500ml of Milli Q H₂O (ELGA LA620, UK), boiled and left to cool. Then 60 µg/ml of ampicillin was added. The LB agar plates containing competent cells were placed overnight in a 37 °C incubator (Luckham R300). The next day plates were

observed for colony growth. Suitable sets of primers were used to screen colonies by means of PCR. Colonies containing the inserted fragment of DNA were taken from the agar plate into 20ml of LB broth medium (Fisher Scientific, UK), to enhance colony growth. The LB medium preparation was carried out by dissolving 12.5g of LB broth in 500ml of Milli Q H₂O, autoclaved, and 60 µg /ml of ampicillin was added, followed by bacterial colony addition. The 50ml tubes containing the LB medium plus colonies were left overnight at 37 °C in a shaking incubator (Gallenkamp Orbital incubator).

2.12 Plasmid DNA Purification, Mini preparation

Mini preps were carried out on transformed bacterial cells grown in 20ml of LB media with appropriate antibiotic in a shaking incubator at 37 °C, and were spun at 5000rpm at 4 °C for 10 min. The pellet was used for DNA preparation using a Qiagen plasmid Mini kit following the manufacturer's protocol. The purity and the amount of the isolated DNA were determined as described earlier in Table 2.1.

2.13 Glycerol Stocks

A positive single colony was grown overnight in LB media with a suitable antibiotic, and 900 µl taken into a 1.5ml Eppendorf tube containing 100µl of glycerol (Sigma, UK). The glycerol stocks were stored at -80 °C.

2.14 DNA Sequencing

For sequencing, 5 µl of plasmid DNA, at a concentration of 100ng/ µl was enclosed in a 0.5ml Eppendorf tube and sent to Source Bioscience Life Sciences. Source Bioscience Life Sciences provided common primers for sequencing such as those that bind to T7 or Sp6 regions. However, for sequencing with custom made primers, 5 µl at 3.2pmol/ µl of each primer per reaction were sent separately with the DNA samples to be sequenced. Sequencing results were analyzed using the Sequence Scanner software v1.0 (Applied Biosystems, UK).

2.15 AT Cloning

For cloning into a vector the 3' or 5' ends of a PCR product must be compatible with vector ends. The taq polymerase enzyme generates 3' end overhang PCR products

which can be ligated into a linearized 3' overhang vector without any modification. Therefore, PCR products were cloned into a linearized pGEM-T vector (Promega, UK) having a "T" overhang at the 3' end, and cloning was conducted according to the pGEM®T easy vector system manufacturer's instructions. The vector is shown in Figure 2.2. This vector has a *LacZ* gene which is present in the multiple cloning region (MCR). IPTG induces β -galactosidase expression. This helps in selection of blue and white bacterial colonies. Bacterial colonies containing pGEM®T vector, when grown on agar plates containing both IPTG and X-gal, start expressing β -galactosidase and will be blue in colour. However, when a DNA fragment is inserted into the MCR, the *LacZ* gene is interrupted and bacterial colonies containing the vector and insert will be visible as white colonies. LB agar plates with 100 μ g/ml ampicillin were further plated with a 100 μ l of 100mM IPTG (Promega, UK) and 20 μ l of 50 μ g/ml X-Gal (Promega, UK). After 30 minutes, plates were transformed with bacterial cells. Colonies were grown overnight at 37°C, and the white colonies were screened using PCR. Positive PCR screened colonies were further grown overnight in LB broth containing ampicillin and the plasmids were then extracted. The purified, extracted, plasmids were quantified and were then sent to Source Bioscience for sequencing. The primers, T7 forward and Sp6 reverse were used to conduct this sequencing.

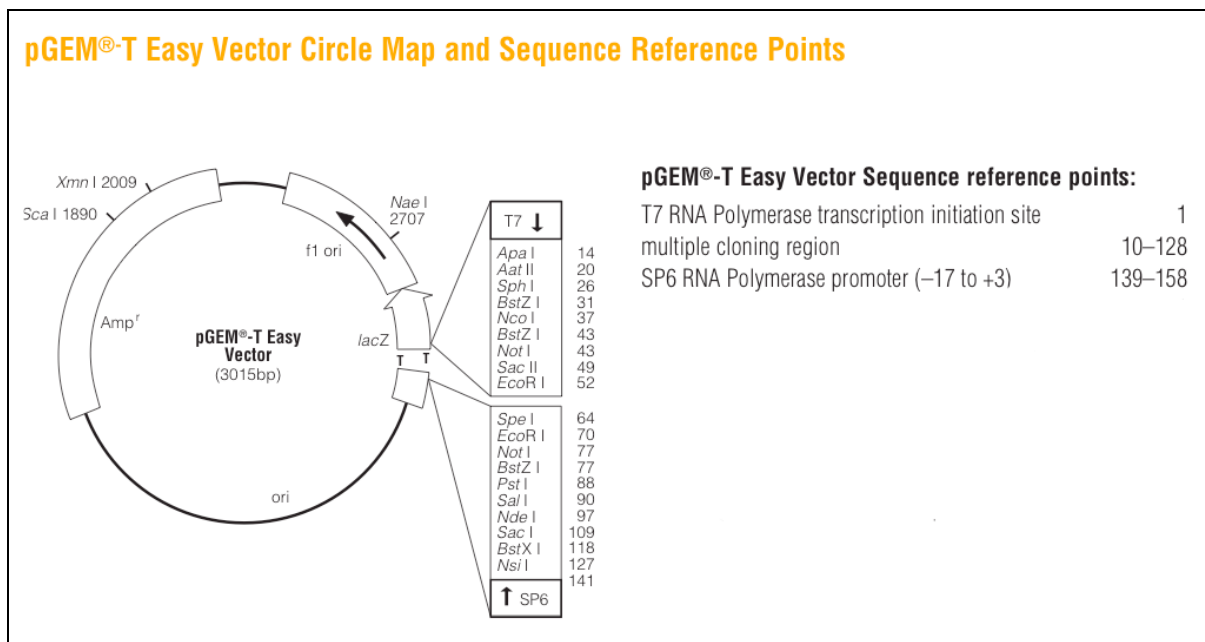


Figure 2.2 Map of the pGEM®T Easy Vector system. The diagram was obtained from the Promega website.

The pGEM®T easy vector system was used in this study for the purposes of AT cloning. As the Map indicates, the vector has a “T” overhang within the MCR in which the DNA fragment of interest, having an “A” overhang, can be inserted.

2.16 Tadpole Anaesthesia

Tadpoles were given anaesthesia before they were processed for DNA/RNA or protein extraction. Tadpoles were placed in water containing 0.025% (w/v) “ethyl 3-aminobenzoate methanesulfonate salt” (Tricaine/MS222; Sigma), (Reed 2005).

2.16.1 Preparation of Brain Tissue Lysates and Protein Assay

For Western Blotting the sample was first run on SDS gel. The protein of interest should first be released from cells and tissues. The cell and tissue lysis is carried out in lysis buffer. For Western blot analysis, frog’s brain and whole tadpoles were used. 300 µl of lysis buffer (Recipe of lysis buffer is shown in Table 2.5) were added per 5 mg of brain tissue. Brain tissue or anaesthetized tadpoles were homogenized manually in a homogenizer in lysis buffer to make a tissue lysate. The tissue lysate was incubated on a roller for 30 minutes at 4 °C for complete lysis. Insoluble material was removed by centrifugation at 14000 rpm for 30 minutes at 4°C. The supernatant was removed to a fresh tube and protein concentration determined with the BCA protein assay kit (Pierce Chemical Co).

Solution name	Components	Amount of chemicals	Storage of Buffers
Lysis Buffer (500ml)	50mM tris base (Sigma, UK)		Once the protease inhibitor tablet was added the solution was stored at -20°C
	(Adjust pH to 8)	3.02g	
	1% triton X-100	0.5g(500µl)	
	150mM NaCl (Sigma, UK)	11.25g	
	5mM EDTA (Sigma, UK)	0.93g	
	one tablet of complete, mini, EDTA-free protease inhibitor cocktail tablets (Roche, UK) was used for every 10ml of buffer		
5X Sample Loading Buffer (10ml)	300mM tris base (Sigma, UK)	4ml	Stored at 4°C
	(Adjust pH to 6.8)		
	SDS 10%	1g	
	Glycerol 0.015%	4ml	
	DTT(dithiothreitol) (Sigma, UK)	0.771g	
	0.015% Bromophenol blue (BioRad)	0.01g	
	dH ₂ O used to make final volume		

Table 2.5 Recipes of lysis buffer and the sample loading buffer.

2.17 Western Blot Sample Preparation

After tissue lysis and protein assay, the protein samples were prepared at a concentration of 1µg/µl in 5x sample loading buffer (recipe given in table 2.5). Samples were boiled for 5 minutes and used immediately or stored at –20°C.

2.18 Peptide Selection and Antibody Production

Peptide regions were selected by following important rules, given in Table 2.6 by using “protein analysis tool/peptide antigen design tool for antibodies”. The peptides were chemically synthesized, injected in rabbit to raise anti-Parapinopsin antibodies and used during Western Blot analysis.

Peptide Characteristics	Explanation						
Peptides sequence for Parapinopsin	N-terminus peptide: “KYRQLRHPINYS” C-terminus peptide: “GRNPWAAEKSSS”						
Homology	Unique amino acid sequences were selected that are highly specific to target protein sequence.						
Ideal protein regions for epitope selection	Antigenic epitopes should be hydrophilic, present at the surface of protein and flexible e.g. loop region. Both Parapinopsin N and C termini peptide regions reside in loop regions, shown in figure 5.2 (Chapter:5) (a) (predicted Parapinopsin secondary structure) and 5.3 (b) (Rhodopsin predicted structure).						
Peptide length	Ideal amino acid length for peptide sequence is “8-20”. Both N and C terminal Parapinopsin peptides were 12 amino acids.						
Peptide solubility	The content of hydrophobic amino acids should be less than 50% and the presence of one charged amino acid for every five amino acids is important.						
Amino acid Sequence composition for both N and C terminus peptides	<table> <tr> <td></td><td>C-terminus peptide</td></tr> <tr> <td>N-terminus peptide</td><td>Neutral: 50%</td></tr> <tr> <td>Neutral: 33.33%</td><td>Acidic: 8.33%,</td></tr> </table>		C-terminus peptide	N-terminus peptide	Neutral: 50%	Neutral: 33.33%	Acidic: 8.33%,
	C-terminus peptide						
N-terminus peptide	Neutral: 50%						
Neutral: 33.33%	Acidic: 8.33%,						

	acidic/basic: 33.33%	Basic: 16.67%
	Hydrophobic: 33.33%	Hydrophobic: 25%

Table 2.6 N and C terminus Peptide selection criteria
(<http://www.lifetein.com/peptide-antigen-design.html>)

2.19 SDS-polyacrylamide Gel Electrophoresis

A BioRad mini apparatus was used to cast two, side by side, Polyacrylamide gels. 15% SDS polyacrylamide gels (7.4 ml of 1M tris pH 8.8, 200µl of 10% SDS, 10ml of acrylamide, 13.2µl of TEMED, 2.5ml of dH₂O and 132 µl of freshly prepared 10% APS; all chemicals were from (Sigma UK) were used to resolve the brain proteins along with 5% stacking gel (1.25ml of 1M Tris pH 6.8, 100µl of 10% SDS, 1.67ml of acrylamide, 20µl of TEMED, 6.8ml of dH₂O and 100µl of 10% APS). Protein sample aliquots were prepared at a concentration of 20µg and loaded onto the gel along with 5µl prestained molecular weight markers (BioRad, UK). Gels were run in duplicate: one for staining and one for blotting purposes. Proteins were separated at 100V for 2 to 3 hours using SDS running buffer (recipe given in Table 2.7).

When the dye reached the bottom of the gel cassette during electrophoresis, the gel was removed. One of the gels was taken out from between the glass plates and placed in a staining solution (0.5% Coomassie Brilliant Blue, 10% glacial acetic acid and 45% methanol) overnight on shaker. Acetic acid helps Coomassie dye to bind with proteins. The rest of the dye was washed away by placing the gel in destaining solution (7.5% glacial acetic acid, and 5% methanol) on a shaker overnight.

For blotting purpose, the gel was given 2 to 3 washes in transfer buffer. In transfer buffer the separated proteins on the gel were transferred to a Hybond ECL nitrocellulose membranes (GE Healthcare) either overnight at 15V or for 1 hour at 100V by using the BioRad “wet transfer” electrotransfer apparatus.

2.20 Western Blot Analysis

Most of the time the following Western Blot protocol was performed. Once the proteins had been transferred to the membrane, the membrane was blocked with 5% milk in PBS (recipe given in table 2.7) for 1 hour on the roller in a 50 ml tube. All the following steps were carried out in a 50ml tube on a roller. 2% milk in PBS-T (recipe given in table 2.7) was prepared for primary anti-frog antibody incubation either 1 hour at room temperature or 4°C overnight on a roller. Membranes were washed 3 times for 10 minutes with PBS-T, followed by secondary antibody incubation i.e. anti-rabbit HRP conjugate, by making a dilution of 1:4000 in 2% milk in PBS-T for 1 hour. Finally the membranes were washed for further 10 minutes washes (3 times) with PBS-T before the proteins were detected with ECL (Enhanced Chemiluminescence) or ECL PLUS reagents (GE Healthcare), and the membranes exposed to X-ray film (GE healthcare).

Solution name	Components	Weight of chemicals in grams	Storage of Buffers
1X Running Buffer (1L)	25mM Tris base (Sigma, UK)	3.03g	Placed at room temperature
	0.2M Glycine (Sigma, UK)	14.42g	
	0.1% (W/V) SDS	1g	
	dH ₂ O	1L	
1X Transfer Buffer (1L)	0.2M Glycine (Sigma, UK)	3.03g	Placed at room temperature
	0.2M Glycine (Sigma, UK)	14.42g	
	20% MetOH	200ml	
	dH ₂ O	800ml	
1X PBS (Phosphate Buffered Saline) (1L)	1 tablets (Oxoid, Basingstoke UK) containing: KCl (0.20g/l), KH ₂ PO ₄ (0.20g/l), NaCl (8g/l), Na ₂ HPO ₄ (1.15g/l),	10 tablets were dissolved in 1L of dH ₂ O and autoclaved	Placed at room temperature
1X PBS-T (Tween)	PBS	1L	Placed at room temperature
	tween	1ml	
Blocking buffers	5% w/v Blocking Buffer	5g	Placed at room temperature
	Skimmed milk powder (Marvel)		
	PBS	10ml	
	2% w/v Blocking Buffer	2g	
	Skimmed milk powder		
	PBS-T	10ml	

Table 2.7 Recipes of the Buffers used during SDS-PAGE electrophoresis and Western Blotting

2.21 Petri Dish Culture of Frog's/ toads Embryos

Early stage tadpoles e.g. stage 20-25 of *X. laevis* and *R. temporaria* were grown in a petri dish in 0.1x MBS (Modified Barth's saline). The recipe of MBS is shown below in table 2.8. MBS media was autoclaved and the antibiotic gentamycin (50 µg/ml) was added in the medium to inhibit bacteria. Plenty of media was added to the petri dish as a lot of media evaporated. Media was changed on a daily basis and dead embryos if any were removed.

Reagents	Weight (g)
NaCl	5.143
KCl	0.075
MgSO ₄	0.120
HEPES	1.192
NaHCO ₃	0.210
CaCl ₂	0.103
dH ₂ O	To 1L

Table 2.8 : Modified Barth's saline (MBS)

The table shows the composition of 1X MBS media for tadpole's growth. (pH 7.8). 100ml of 1X MBS media were taken and diluted to 900ml with autoclaved water to make 0.1X MBS and used for the tadpole's growth. (All chemicals used from Sigma UK)

2.22 Culturing of free Swimming Tadpoles

On reaching stage 43, embryos became free swimming tadpoles. At this stage tadpoles were moved to a larger size beaker filled with 1X Modified Frog Ringer's (MR). Recipe of 10X MR is given in table 2.9.

Reagents	Weight (g)
NaCl	58.44 g
KCL	1.34 g
CaCl ₂	2.22 g
MgCl ₂	2.03 g
Hepes	11.92 g
dH ₂ O	1L

Table2.9 : MR (Modified Frog Ringer's) solution composition

The table shows the composition of 10X (1L) MR solution. To make 1X MR, 100ml of 10X was added in 900ml of dH₂O. (All chemicals used were from Sigma UK)

2.23 Tadpoles Preservation

Tadpoles were preserved in liquid nitrogen or in 75% ethyl alcohol and stored at -80 °C.

2.24 Cell Culture Work

A *X. laevis* melanophore cell line (kind gift from Dr. David Sugden) was used during the project to study Parapinopsin mRNA through Northern blotting. Originally the melanophores came from a *X. laevis* embryo. Melanophore cells were grown in Leibovitz L15 medium (Invitrogen life technologies). The medium was supplemented with 15% v/v foetal calf serum, 100U/ml penicillin and 0.1mg/ml streptomycin (all Invitrogen Life Technologies). As amphibians are poikilothermic *i.e.* body temperature fluctuates according to surrounding environment, there is no need of a particular temperature and exchange of gases for melanophores. The melanophores were grown at room temperature between 20-27°C.

The melanophore cells were grown and fed twice a week in 175cm² culture flasks (T175, Nalge Nunc International). After approximately a week, the melanophores became confluent and then passaged /sub cultured. Before sub-culturing and passaging the melanophores were washed with 10-15 ml of 0.7x PBS and then the flasks were treated with 5ml of trypsin (Invitrogen life technologies) in 0.7x PBS to dissociate the

melanophore cells, as melanophore cells are adherent to the flask surface. After trypsinization, the cell lines were provided with 10ml of growth medium, and then spun at 400xg/ 2 min before harvesting.

2.24 RNA Sample Preparation and RNA Denatured Gels

Northern Blotting is a molecular biology technique in which gene expression is studied by the presence of RNA in the sample. During the project, Northern Blotting started with total mRNA extraction from *X. laevis* melanophores. Total mRNA was extracted using oligo dT and its absorbance measured by following the protocols mentioned under the headings of 2.1 and 2.2. The total, extracted, melanophore mRNA samples were separated on the basis of size by agarose gel electrophoresis (Recipe of gel shown in Table 2.10 below) and transferred to nylon membrane by capillary action.

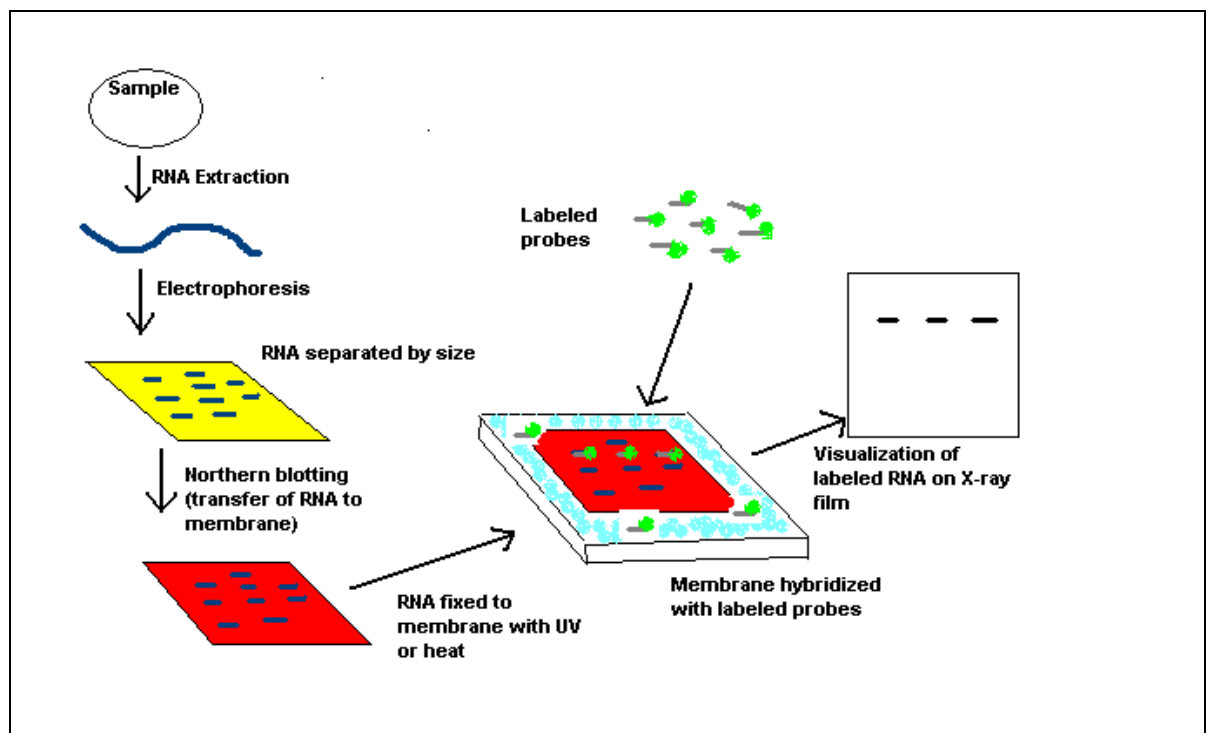


Figure 2.3 Schematic presentation of the Northern Blotting mechanism

The diagram shows the different steps in Northern Blotting

(http://en.wikipedia.org/wiki/File:Northern_Blot_Scheme.PNG)

5 µl of RNA (10-20 µg) and RNA standards were mixed with 5 µl of RNA gel loading buffer (Recipe of loading buffer given in table 2.9 below) and denatured at 65 °C for 15 minutes then chilled immediately on ice. Samples were loaded on 1% agarose gels (0.5g of agarose weighed in DEPC treated flask, 43.25ml DEPC treated water then added and 5ml of 10X MOPS-EDTA; recipe shown in table 2.10, gel melted in oven and loaded in DEPC treated gel tank; Max fill, Fisher brand, UK). A voltage of 50V (110A) was applied across the gels for 2-3 hours or when the dye reached the bottom of gel, and the gel observed under UV. Recipe of running buffer 1X MOPS-EDTA is shown in Table 2.9. The gel was melted in an oven and loaded into a DEPC treated tank; Max fill (Fisher Brand UK).

For blotting purposes, the gel was soaked in transfer buffer; 10X SSC (Sigma UK, 20X SSC buffer). In transfer buffer (between a sandwich of sponge and filter papers) the separated RNAs from a gel were transferred to a Hybond ECL nitrocellulose membrane (GE Healthcare) for 1 hour at 100V by using the BioRad “wet transfer” electrotransfer apparatus. The RNA was cross-linked to the membrane in a UV cross linker (UV stratalinker™ 1800).

Buffer name	Components	Storage information
10X MOPS-EDTA buffer (1L)	0.2M MOPS = 41.86g	Autoclave the buffer and make final volume up to 1L by adding DEPC treated H ₂ O.
	50mM Sodium Acetate Anhydrous = 4.102g 10mM EDTA = (20ml of 0.5M solution pH8.0) dH ₂ O = 800ml	
RNA gel loading buffer	Formamide = 750 µl	Aliquoted and store at -70 °C
	37% Formaldehyde = 250 µl	
	10X MOPS-EDTA buffer = 150 µl	
	86% Glycerol = 116µl	
	0.25% Bromophenol Blue = 122 µl	
Running Buffer	0.25% Xylene Cyanol = 122 µl	Prepared fresh from 10X MOPS-EDTA
	Ethidium Bromide(10mg/ml) = 6 µl	
1X MOPS-EDTA (1L)	10X MOPS-EDTA = 100ml DEPC H ₂ O = 900ml	

Table2.10 Buffer Recipes for RNA gels

The table shows the recipes of RNA gels and storage information. All chemicals were from Sigma UK.

2.25 Northern Blotting

Northern blotting was done by using DIG Northern starter kit (Cat. No.12039672910).

The protocol was followed according to the manufacturer's instruction.

2.26 Protein Expression studies

2.26.1 Parapinopsin Gene Synthesis

Although GPCR proteins are difficult to express in a microbial system like *E.coli*, we have chosen to use a GPCR expression system developed in Prof Georgiou's laboratory, Dept. of Chemical Engineering UT Austin Texas, USA. Link *et al.*, 2008; Skretas & Georgiou, 2008 and Skretas, 2012 used a fusion protein construct consisting of GFP (green fluorescent protein) fused to the GPCR and expressed the protein in *E.coli*. GFP fluorescence hence acts as a reporter to monitor for expression of the GPCR protein. Link *et al.*, 2008 also revealed an increase in expression of GPCR when MP was co-expressed with membrane-anchored AAA+ protease FtsH. We adopted the same strategy in order to overexpress the Parapinopsin protein in *E.coli*.

A codon optimized Parapinopsin gene with a start and a stop codon was engineered into pUC57 Kan (Kanamycin) by GENEWIZ, USA. The Parapinopsin gene's 5' and 3' ends were modified in line with the expression vector, discussed below.

2.26.2 Expression Vectors

The anhydrotetracycline-inducible plasmid pASKGPCR (modified version of pASK75) expression vectors (kind gift from Dr. George Georgiou, Dept. of Chemical Engineering, UT Austin, Texas USA) were used during the current studies. Two types of pASKGPCR constructs were used, one with GFP, and the other without GFP. The figures of pASKGPCRs are shown in Figure 2.4 a & b.

2.26.3 Primers used during Expression Studies

Following is a list of PCR primers in Table 2.11 used during the engineering of the constructs to be employed in the Protein expression studies. BLAST searches looking for similarities in Primer sequences were carried out by comparing selected primer sequences to the sequences present in the database using Primer-BLAST at NCBI. Primers were made by IDT (Integrated DNA technologies, UK).

araP-PCR Primer	Sequence of primer set	Annealing Temperature	Product size (bp)
ParaP1 Forward	Forward5'dtccttctagatagcgatcaagaataacatggattac aaagacgacgatgacaagATGGCCGACGAGGCCCTGT 3'	66°C	1131
ParaP2 Forward	Forward5'dgcgccttgatcgctatctagaaggaggtcgtcgtctt tgtaatccatgttattaacctccttctagatagcgatc3'	65°C	1188
Common reverse primer pnpR	Reverse5'dtggccaggtggccccggcactgcagcatcatcaccatc accactagaagctt3'	62°C	1131/1188
pASK75_1	Forward5'dcatcgaatggccagatgat3' Reverse5'dgtctgccgtttaccgctact 3'	59°C	1549
pASK75_3	Forward5'dcaaagacgacgatgacaag3' Reverse5'dgcagcatggaaccagtat3'	58°C	1014

Table 2.11 Primer sets used in expression studies, 5'→3'

The table above shows the primer name in 1st column, its sequence in the 2nd column, the annealing temperature used for PCR in the 3rd column and the expected product size (bp) shown in column 4. Adaptor sequences at the 5' end of ParaP1 Forward Primer: Black text nucleotides represent the RBS sequence, red text nucleotides represent the FLAG sequence and blue highlighted is the sequence match with the Parapinopsin gene. ParaP2 Forward primer: yellow highlighted sequence represents the *Xba*I restriction enzyme sequence. Adaptor sequences at the 3' end of ParaP reverse common primer: Blue highlighted sequence matches with gene are illustrated, in red for restriction enzyme *Pst*I, in grey for His₆, in yellow for stop codon and in pink for restriction enzyme *Hind*III.

2.26.4 Cloning of the Parapinopsin gene in pASKGPCR

The codon optimized Parapinopsin gene was cloned in pASKGPCR by adding *Xba*I restriction site, SD (ribosomal binding site), FLAG (immunological detection) nucleotide sequences at 5' end by a two steps PCR. The 3' end of the Parapinopsin was modified by adding a *Hind*III restriction site and octaHis nucleotide sequences by a single step PCR. Both plasmid pASKGPCR and Parapinopsin gene were digested with *Xba*I and *Hind*III restriction enzymes (New England Biolabs, UK), and ligation reaction was carried out by T4 DNA Ligase (Promega USA).

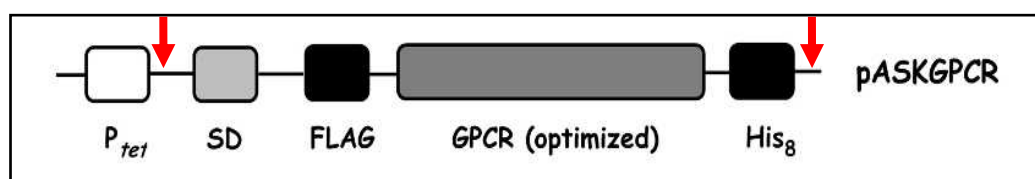


Figure 2.4 (a) Expression vector pASKGPCR (A. JAMES LINK *et al.*, 2008)

The figure above shows the expression cassette of pASKGPCR plasmid where the protein expression is under the control of anhydrotetracycline (Ptet), SD is the Shine Dalgarno sequence where the ribosomes bind. A GPCR (G-coupled protein receptor) codon optimized gene “Parapinopsin” is present between “FLAG” N-terminal tag for immunological detection and C-terminal octaHis histidine tag for purification by immobilized metal affinity chromatography. Red arrows show the restriction sites for *Xba*I before SD sequence and *Hind*III after the His tag.

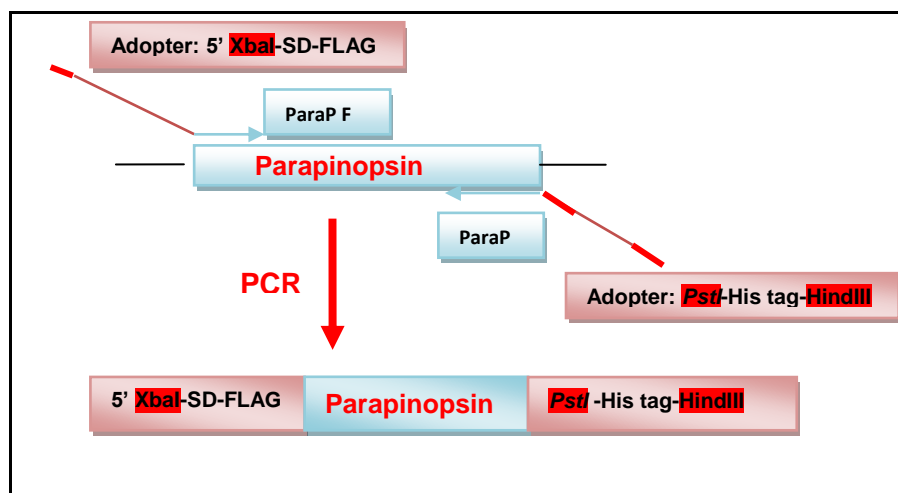


Figure2.4 (b) The 5' and 3' modification of the Parapinopsin gene

The figure above shows the addition of 111 nucleotides “adopter” sequence at 5' end of Parapinopsin gene by two steps PCR. In step 1 of PCR, in 5' end of ParaP forward primer (~20bbp), 65 nucleotides were added which act as template for 2nd step PCR to add rest of nucleotides. At 3' end of gene the “adopter” sequence of two restriction sites and His tag was added at 5' end of ParaP reverse primer.

2.26.5 pASKGPCR-GFP

The codon optimized Parapinopsin gene (Figure 2.4b) was cloned in pASKGPCR-GFP by digesting both plasmid and Parapinopsin with restriction enzymes *XbaI* (at 5' end) and *PstI* (at 3' end (New England Biolabs, UK). The ligation reaction was carried out by T4 DNA Ligase (Promega USA).

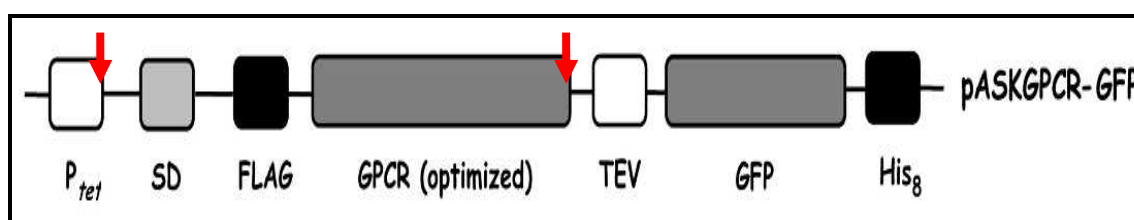


Figure2.5 Expression vector pASKGPCR- GFP (A. JAMES LINK *et al.*, 2008)

The figure above shows expression cassette of pASKGPCR-GFP plasmid, which encodes the C terminal GFP (green fluorescent protein) and the sequence for cleavage site “TEV” (tobacco etch virus) and His tag. Red arrows show the restriction sites for *XbaI* before SD sequence and *PstI* after GPCR (G-coupled protein receptor), where Parapinopsin gene was cloned by digesting with same restriction enzymes.

2.26.6 Co- Expression of pASKGPCR and FtsH

The cloned Parapinopsin gene in pASKGPCR was also co-expressed with FtsH protein. The AAA+ protease FtsH protein is cloned in arabinose inducible pBAD33 vector

(Figure2.6) by Dr. George Georgiou, Texas USA and his team. Membrane-anchored AAA+ protease FtsH protein increased the expression of GPCR (Link *et al.*, 2008). During the current studies both plasmids pASKGPCR (containing Parapainopsin gene) and pBAD33 (containing FtsH coding sequence), were co-transformed in MC4100A *ara*⁺ *E. coli* strain (kind gift from Dr. Georgios Skretas, Athens Greece) to obtain maximum Parapainopsin protein expression.

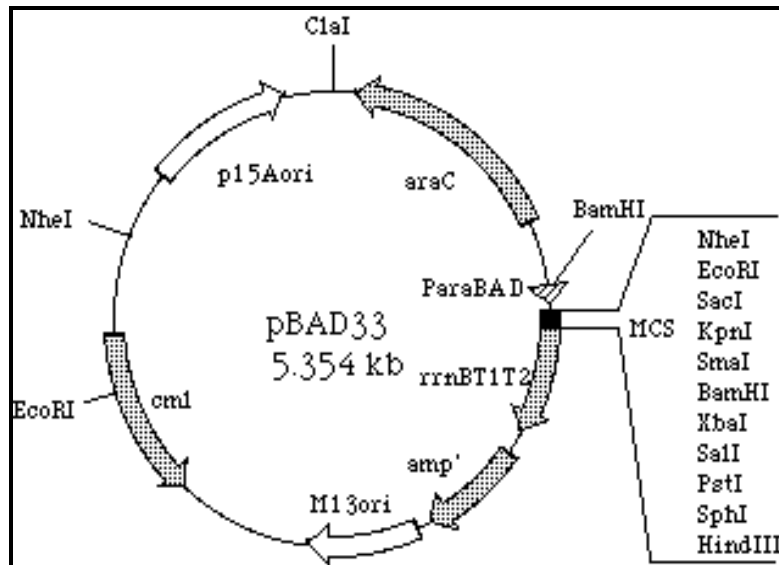


Figure2.6 pBAD33 vector

The figure above shows the map of arabinose inducible pBAD33 plasmid. The plasmid was digested with *SacI* and *HindIII* restriction enzymes where the FtsH protein was cloned by (A. James Link *et al.*, 2008) and co-expressed with pASKGPCR plasmid.

(<http://www.shigen.nig.ac.jp/ecoli/strain/cvector/map/pBAD33.gif>)

Chapter 3 Sequencing of Parapinopsin in Frogs

3.1 Aim of Project

This chapter describes the sequencing of Parapinopsin cDNA, and for this the tadpoles of different species of frogs *X. laevis*, *R. temporaria* and *R. tigrina* were used. The only reported Parapinopsin sequence at NCBI is that of *Xenopus tropicalis* (NM_213665.1). The *X. tropicalis* sequence is shown in appendix 1. Hence the sequence of *X. tropicalis* Parapinopsin cDNA (1117bp) was taken as the reference sequence; primers were designed against *X. tropicalis* Parapinopsin cDNA sequence shown in chapter 2 Table 2.2 and used against cDNA of different frogs and toads. To study Parapinopsin, first the tadpoles were grown to get their cDNA and genomic DNA.

3.2 Sequencing of *Xenopus tropicalis* Parapinopsin cDNA

The cDNA sequence of *X. tropicalis* Parapinopsin reported at NCBI is 1117bp (NM_213665.1) was taken as Parapinopsin reference sequence. Various regions of the cDNA sequence of *X. tropicalis* were randomly selected to design gene specific PCR primers, shown in the primer Table. Primer set named PNPd were used against gDNA and cDNA (commercially purchased from Portsmouth) of *X. tropicalis* respectively. PCR was performed to confirm whether the sequence of Parapinopsin matched that of the reported sequence.

3.2.1 PCR amplification of Parapinopsin cDNA of *X. tropicalis*

The cDNA of *X. tropicalis* was used as template for PCR amplification by using primer set PNPd (Chapter:2, sequences shown in Table 2.2). A 900 bp PCR product was expected to be produced. The PCR product is shown in Figure 3.1. The PCR product of PNPd of *X. tropicalis* cDNA was then purified, quantified and sent for sequencing at Source Biosciences life Sciences.

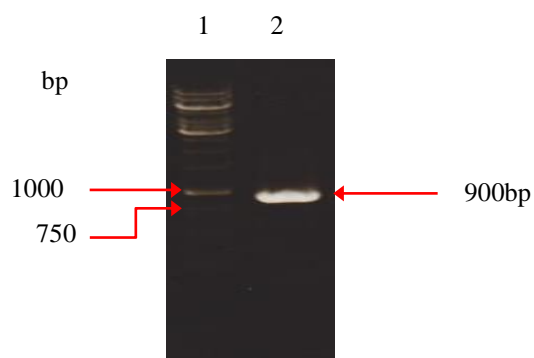


Figure 3.1 Electrophoretic Analysis of PCR reaction using PNPd on *X. tropicalis* cDNA

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2: PCR product of the PNPd primer set

A 1% agarose gel was used to analyse PNPd PCR products through electrophoresis. In lane 2, the 900bp band indicates that the cDNA PCR reaction was successful.

Purified PCR products using primer set PNPd for *X. tropicalis* cDNA were quantified and also sent for sequencing at Source Biosciences life Sciences. Sequencing results showed 100% similarity with the reported sequence of Parapinopsin for *X. tropicalis*. The next step was the cDNA study of *X. laevis* and *R. temporaria* Parapinopsin.

3.3 Sequencing of *Xenopus Laevis* Parapinopsin cDNA

3.3.1 PCR Amplification of *Xenopus Laevis* Parapinopsin cDNA

Total RNA was extracted from *X. laevis* whole tadpoles of stage 53. Extracted RNA was treated with DNase to remove possible DNA contamination. After RNA quantification, a reverse transcriptase reaction was carried out. To see whether cDNA synthesis was successful with no contamination of gDNA, a set of primers was designed against housekeeping gene β -actin. Then a PCR reaction was done on the synthesized cDNA of *X. laevis* using the β -actin primers (found in Table 2.2 in chapter Material and Methods) together with negative control PCR without template. The PCR reaction results are shown in Figure 3.2 and indicate that the cDNA synthesis was successful.

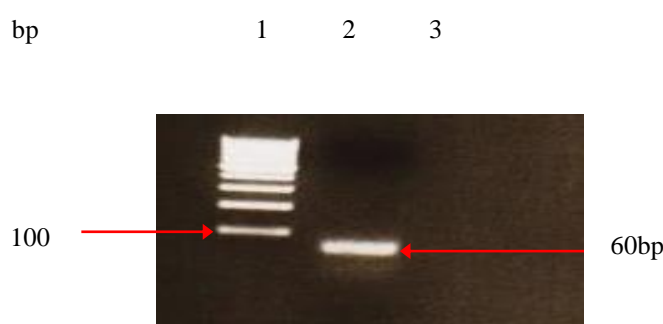


Figure 3.2 Electrophoretic Analysis of PCR reaction using β -actin primers on *X. laevis* cDNA.

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR product of the β -actin primer

Lane 3: Negative control; PCR without template

A 2% agarose gel was used to analyze β actin PCR products through electrophoresis. In lane 2, the 60bp band indicates that the RNA purification and cDNA synthesis of *X. laevis* tadpoles was successful. Lane 3 is the negative control PCR without template.

A PCR reaction was conducted by using primer sets PNP1 and PNPd (sequences shown in Table 2.2). The expected PCR product sizes of PNP1 and PNPd are 320bp and 900 bp respectively. The electrophoretic analysis is shown in Figure 3.3. The PCR products of PNP1 and PNPd *X. laevis* cDNA were then purified, quantified, and 500ng was used for AT cloning. The remaining PNP1 and PNPd PCR products were stored at -20°C.

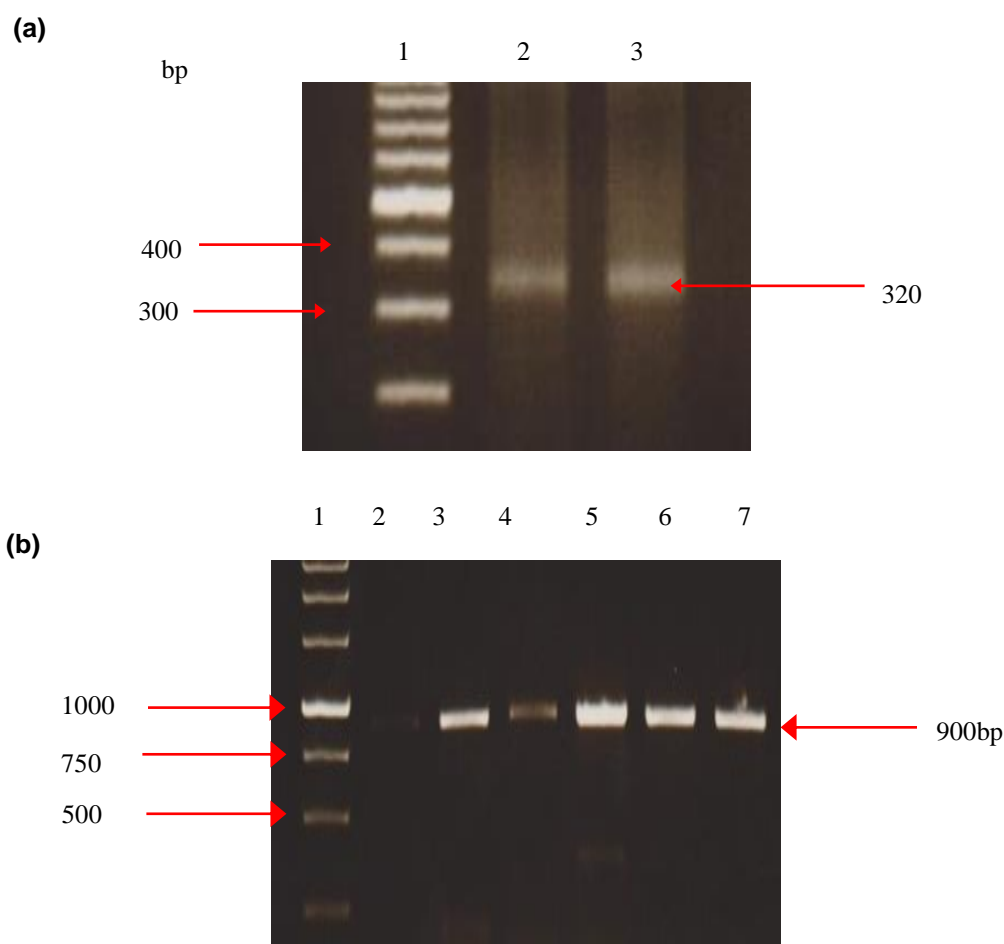


Figure 3.3 Electrophoretic Analysis of PCR reaction using PNP1 (a) and PNPd (b) primer sets on *X. laevis* cDNA.

(a)

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR product of PNP1 on *X. laevis* cDNA

Lane 3: Repeat of lane 2

A 2% agarose gel was used to analyse the PNP1 PCR products through electrophoresis. In lane 2, the 320bp band indicates that the cDNA of *X. laevis* tadpoles contains Parapinopsin. Lane 3 is the repeat of lane 2.

(b)

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2: PCR reaction did not work

Lane 3-7: gradient PCR products of PNPd on *X. laevis* cDNA

A 1% agarose gel was used to analyse the PNPd PCR products through electrophoresis. In lanes 2-7, the gradient PCR reaction between 55°C- 65 °C was carried out. PCR did not work in lane 2. The 900bp band in lanes 3-7 indicates that the cDNA of *X. laevis* tadpoles contains Parapinopsin gene.

3.3.2 AT Cloning of the PNP PCR Products in the pGEM® T Easy Vector System

The amplified and purified PNP1 (320bp) and PNPd (900bp) PCR products were cloned in the pGEM®T easy vector system. This was done through the process of AT cloning. The PCR products have an “A” overhang, and the pGEM®T easy vector has a complimentary “T” overhang in its multiple cloning region (MCR) for easy insertion of the PCR products.

The ligation reaction was done overnight at 4°C. The reactions were analysed by PCR using T7 forward and SP6 reverse primers or T7 forward and PNP1 reverse primers, where bands of ~500bp and ~400bp respectively would be expected in case of pGEM PNP1. Whereas PCR of the ligation reaction with T7 forward and SP6 reverse primer and T7 forward and pGEM PNPd reverse primer would give as ~1000bp PCR product in the case of PNPd. The results of the PCR can be seen below in Figure 3.4 a, b & c.

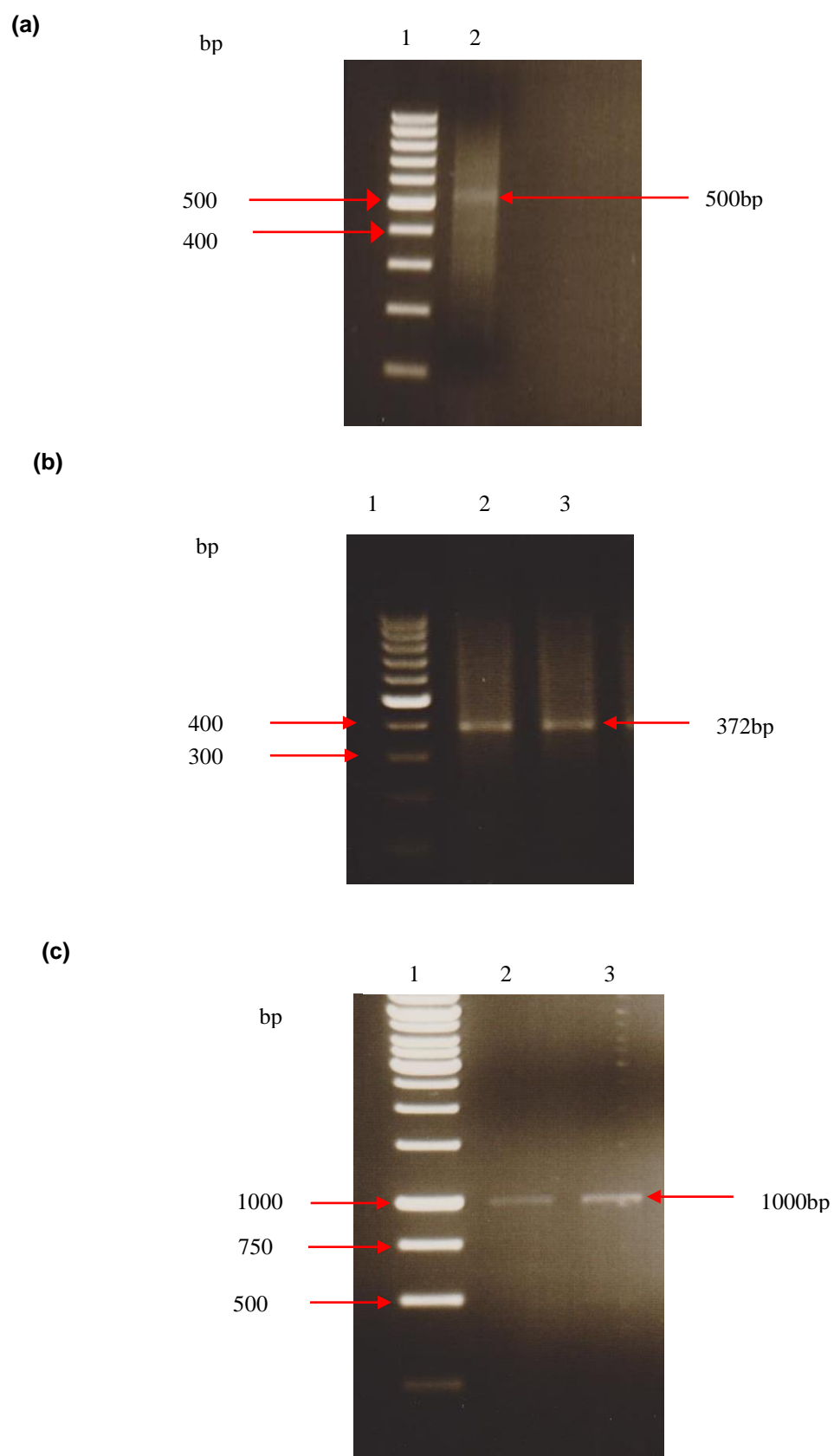


Figure 3.4 Electrophoretic Analysis of ligation reaction PCR

(a)

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR product of T7 forward and Sp6 reverse primers with pGEM PNP1 ligation reaction

(b)

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR product of T7 forward and PNP1 reverse primers with pGEM PNP1 ligation reaction

Lane 3: PCR product of T7 forward and PNP1 reverse primers with pGEM PNP1 in a repeat ligation reaction

(c)

Lane 1: GeneRuler™ 1000bp DNA Ladder

Lane 2: PCR product of T7 forward and Sp6 reverse with pGEM PNPd ligation reaction

Lane 3: PCR product of T7 forward and Sp6 reverse with pGEM PNPd in a repeat ligation reaction

A 2% agarose gel was used to analyse PCR products of ligation mix through electrophoresis. (a) In lane 2, the ~500bp PCR product indicates that the PNP1 PCR product is successfully ligated in pGEM vector. (b) In lanes 2 and 3, the ~400bp of T7 forward and PNP1 reverse PCR product band indicates that successful ligation of PNP1 PCR product in pGEM vector. (c) In lanes 2 and 3, the ~1000bp PCR product band indicates successful ligation of PNPd PCR product in pGEM vector.

Half the amount of both ligation mixtures was used to transform separately 50µl of aliquoted XL1Blue competent bacterial cells. The cells were then plated on agar plates with the appropriate antibiotic and were kept overnight in an incubator at 37°C. The colonies on the plates were screened by means of PCR. PCR was done using T7 forward/ PNP1 reverse (PCR product specific) and T7 forward/ PNPd(PCR product specific) reverse primer. PCR products were analysed by electrophoresis, and the gels are illustrated in the Figure 3.5 a & b.

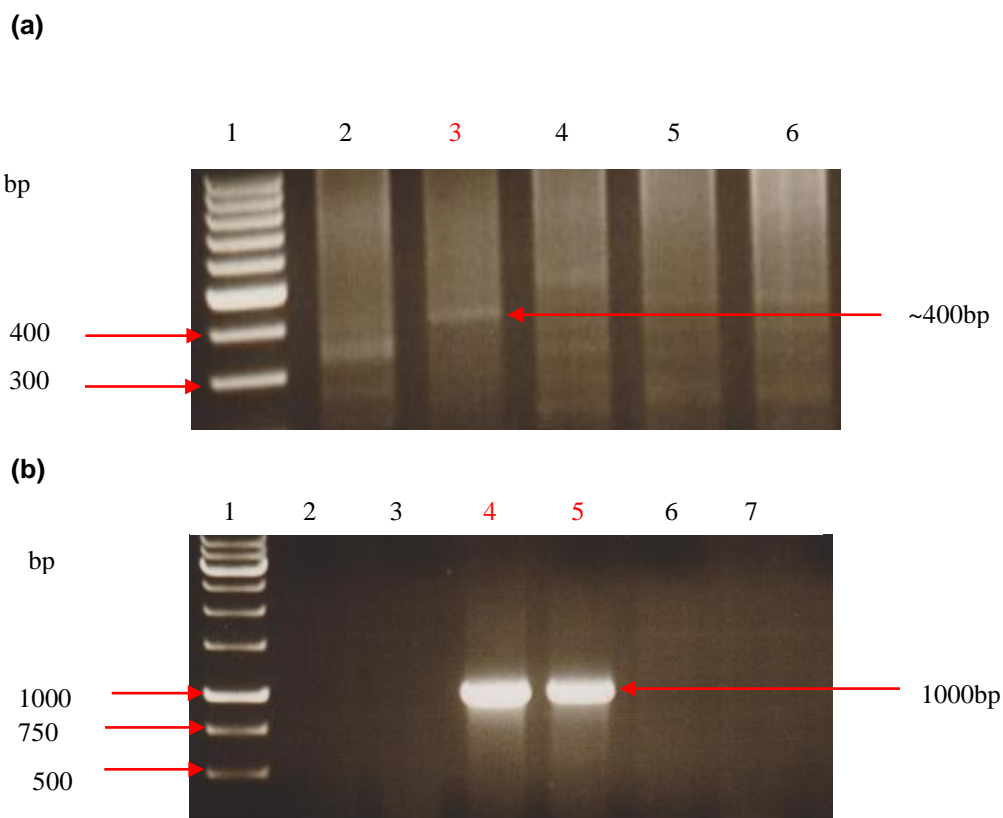


Figure 3.5 Electrophoretic Analysis of the PCR products of the pGEM vector Colony Screen

(a)

Lane 1: GeneRuler™ 100 DNA Ladder

Lanes 2, 4-6: Negative Colony screen PCR

Lane 3: Positive Colony screen PCR

The primer set T7 forward and PNP1 reverse was used for the above colony screen and the expected PCR product size is ~400bp. Lanes 2, 3, 4 and 5 all have negative colony screen PCR products. But lane3 has band size of ~400bp (~52bp for the wild type and 320bp for the PNP1 PCR product). This colony was considered to be a positive colony and highlighted in red in the gel picture. The whole colony was picked up from the agar plate and grown in 20ml of LB for mini prep with the appropriate antibiotic.

(b)

Lane 1: GeneRuler™ 1kb DNA Ladder

Lanes 2, 3, 6 and 7: Negative Colony screen PCR

Lane 4&5: Positive Colony screen PCR

The primer set T7 forward and PNPd reverse was used for the above colony screen and the expected PCR product size is ~1000bp. Lanes 2, 3, 6 and 7 were all negative in the colony screen of the PCR products. But lanes 4 and 5 have a band size of ~1000bp. These colonies were considered to be positive colonies and are

highlighted with red in the gel picture. The colonies were picked up from the agar plate and grown in 20ml of LB for mini prep with the appropriate antibiotic.

3.4.3 Sequence Analysis of the PNP1 products

Two cloned plasmids containing PNP1 (320bp) and PNPd (900bp) PCR products were extracted by means of mini preps. The plasmids were analysed by PCR, first with T7 forward/ PNP1 reverse and with T7 forward/ PNPd primer sets. The PCR products of mini preps were analysed by gel electrophoresis, and are shown in Figure 3.6 a & b. Once, positive colonies were confirmed by PCR, the plasmids were sent to Source Bioscience Life Sciences for sequencing.

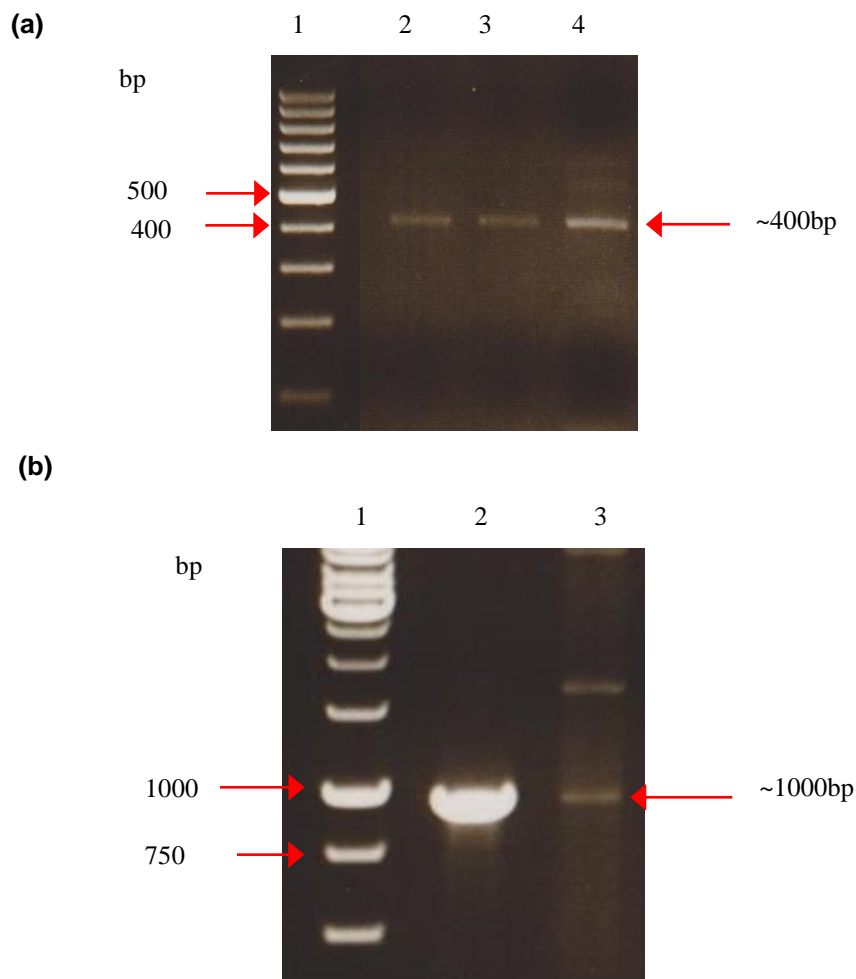


Figure 3.6 Electrophoretic Analysis of the PCR products of (a) cloned pGEM PNP1 (b) pGEM PNPd

(a)

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: Colony 3 PCR product of the pGEM-PNP1 vector using primer set T7 forward with PNP1 reverse

Lanes 3 & 4: Repeat of lane 2

(b)

Lane 1: GeneRuler™ 1kb DNA Ladder

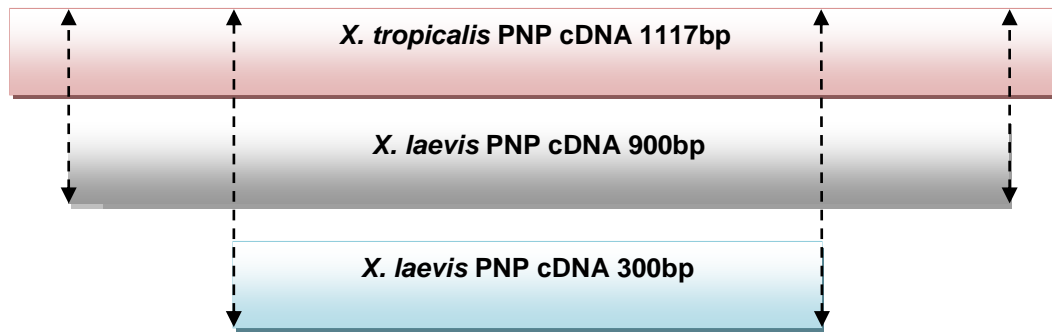
Lane 2: colony 4 PCR product of the pGEM-PNPd vector using primer set T7 forward with PNPd reverse

Lane 3: colony 5 PCR product of the pGEM-PNPd vector using T7 forward with reverse

To confirm that the plasmids had the right insert in the right orientation, the PCR reaction was conducted on pGEM-PNP1 plasmid and pGEM-PNPd plasmid with T7 forward/PNP1 reverse and T7 forward/PNPd reverse respectively. If a PNP1 PCR product has been inserted in pGEM the size of the PCR product expected for the primer T7 forward/PNP1 reverse sets is ~ 400bp and for T7 forward/PNPd reverse, ~1000bp. The electrophoretic analysis illustrates that the resultant PCR products are of the sizes expected, thus confirming that the plasmids had the correct insert.

The sequences of PNP1 (320bp) and PNPd (900bp) of *X. laevis* cDNA were analysed and compared with the sequence of *X. tropicalis* found at the NCBI website, accession numbers NM_213665.1. The sequencing results confirmed that 900bp of the cloned *X. laevis* cDNA used in this study was 100% homologous with the sequence found in the literature for *X. tropicalis* Parapinopsin (except for 162 nucleotides at 5' and 55 nucleotides at the 3' end, highlighted in green, shown in figure 3.7 were not sequenced). Attempts were made to sequence the 5' regions (from 1-162bp) of *X.laevis* Parapinopsin by designing different sets of primers, but PCR reaction failed due to unknown reasons. The sequencing results of PNP1 and PNPd are shown below in Figure 3.7.

a) Cloned *X. laevis* PNP cDNA matches with *X. tropicalis* PNP cDNA sequence



b) Highlighted *X. laevis* PNP sequence and *X. tropicalis* PNP cDNA sequence

tgtgccctgacaatttcggtgactaaaaagctcagaaccaaaggccgacgaggctcttctccaccaatgatgaatgtgaccaatga
 agagatgcaccctgggaagggtctgatgccccggattggttacaccattctggcttaataatggccgtattctgtgccgcccgtctcttctc
 aacgtcacggtgattgtggtgactttcaaatatcgccaactgcgccaccaatcaactactcgctggtaacctggccatcgccgacctt
 ggcgtaaccgtgctgggcggggccttgacagtggagacaaatgccgtgggctactttaacctggggagagtgggctgctcattgagg
 gtttcgctggcatttttggcatcgcggttctgtgcaccattgcagtaatacgcttgaccgagtggtcggtgtgcaagccaatgggca
 cactcacattcaccaccaacaagcactggccggcattgctgcttcttgatctggtcgctcatatggaatacgccccgctgttcggctg
 gggtagctacgagctggaaggggtgatgacatcctgcgccccaaactggtacagcgcgacccagtcacatgtcctacatcgtctgc
 tacttctcttctgcttgccatcccattcctcatcatcgctgggtcctacgggtacctgatgtggactctgcggcaggttgccaagttgggtg
 agcgggaaggaggtacaacgagtaaagccgaggtgcaagtctcccgtatggtgattgtcatgatcttggttctctggtctgttgctcccg
 tacgccgccttgccatgacggttgggcaaataccggaatgcacattgacccattatagccaccgtacccatgtacctgacaaaac
 cagcaccgtctataatccaattatctacatttcatgaacaagcagttccaagaatgtgtcattcccttctgttctgcgggaggaatccttgg
 gctgctgaaaagtcaagttctatggaaactctatcagcgtaacgagtggaacgcctacgaaacgtggccaagtggctccggcgtaac
 gtgctttatctccagtataaccggc

Figure 3.7 *X. laevis* sequence matches with *X. tropicalis* Parapinopsin cDNA sequence

(http://www.ncbi.nlm.nih.gov/nuccore/NM_213665.1)

Part “a” of above Figure depicts the whole 1117bp PNP sequence of *X. tropicalis* reported at NCBI. In *X. laevis*, the 320bp and 900bp sequences match exactly with *X. tropicalis* PNP cDNA; primers bound to these regions were highlighted in yellow. In part “b” the *X. tropicalis* PNP sequence, blue highlighted, is the *X. laevis* 320bp sequence result, while the 900bp sequence is shown in grey. Only grey highlighted 900bp sequence from 163 -1062 nucleotide in *X. tropicalis* and *X. laevis* cDNA showed 100% homology. Primers designed in green highlighted regions from 1-162 & 1062-1117 nucleotides yielded no PCR products.

3.5 PCR amplification of *Rana Temporaria* Parapinopsin cDNA

Total RNA was extracted from *R. temporaria* whole tadpoles of stage 53. Extracted RNA was treated with DNase to remove possible DNA contamination. After RNA quantification, a reverse transcriptase reaction was carried out. Then a PCR reaction was done on the synthesised cDNA using β -actin primer, sequence of primer found in Table 2.2. The PCR reaction results, shown in Figure 3.8, conducted on the synthesised cDNA from the *R. temporaria* tadpoles indicate that cDNA synthesis is successful.

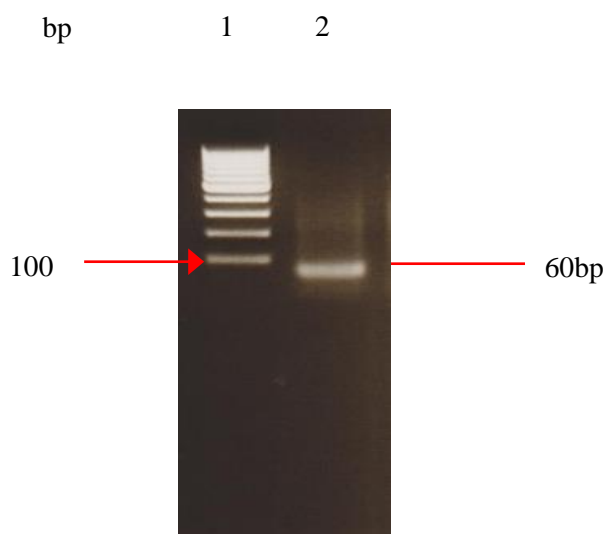


Figure 3.8 Electrophoretic Analysis of PCR reaction using β -actin primers on *R. temporaria* cDNA

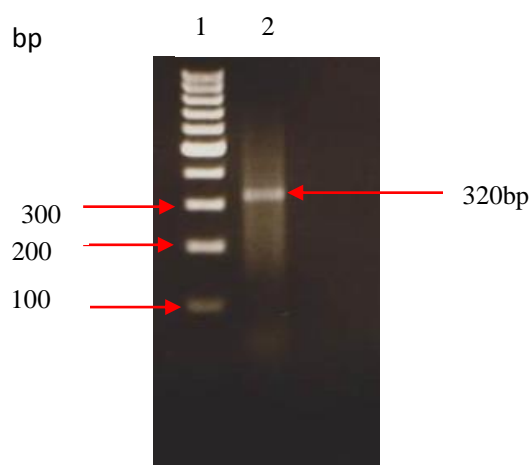
Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR product of the β -actin primer

A 2% agarose gel was used to analyse β -actin PCR products by electrophoresis. In lane 2, the 60bp band indicates that the RNA purification and cDNA synthesis of *R. temporaria* tadpoles was successful.

When the cDNA synthesis of *R. temporaria* tadpoles was confirmed, a PCR reaction was conducted by using primer sets PNP1 and PNPd (sequences shown in Table 2.2). The expected PCR product sizes with PNP1 and PNPd are 320bp and 900 bp respectively. However PCR products corresponding to the whole cDNA of *X. tropicalis* including bases 1-162 were not obtained. The electrophoretic analysis is shown in Figures 3.9. The PCR products of PNP1 and PNPd *X. laevis* cDNA were then purified, quantified, and 500ng was used for AT cloning.

(a)



(b)

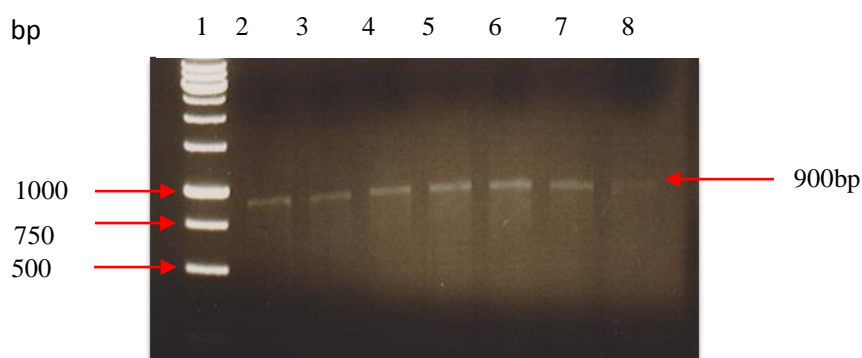


Figure 3.9 Electrophoretic analysis of PCR reaction using PNP1 (a) and PNPd (b) primer sets on *R. temporaria* cDNA.

(a)

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR product of PNP1 from *R. temporaria* cDNA

A 2% agarose gel was used to analyse the PNP1 PCR products through electrophoresis. In lane 2, the 320bp band indicates that the cDNA of *R. temporaria* tadpoles contains Parapinopsin.

(b)

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2-8: PCR product of PNPd from *R. temporaria* cDNA

A 1% agarose gel was used to analyse the PNPd PCR products through electrophoresis. Lanes 2-8, show the results of a gradient PCR reaction between 55°C- 65 °C. The 900bp band in lanes 2-8 indicates that the cDNA of *R. temporaria* tadpoles contains Parapinopsin gene.

3.5.1 AT Cloning of the *R. temporaria* PNP PCR Products in the pGEM® T Easy

Vector System

The amplified PNPd (900bp) PCR product was cloned in the pGEM®T easy vector system. This was done through the process of AT cloning as explained in section 3.4.2.

The ligation reaction was done overnight at 4°C. The reaction was analysed by PCR using either T7 forward and SP6 reverse primers, or T7 forward and PNPd reverse primers. A ~1000bp PCR product is expected with the latter set. The results of PCR can be seen below in Figure 3.10.

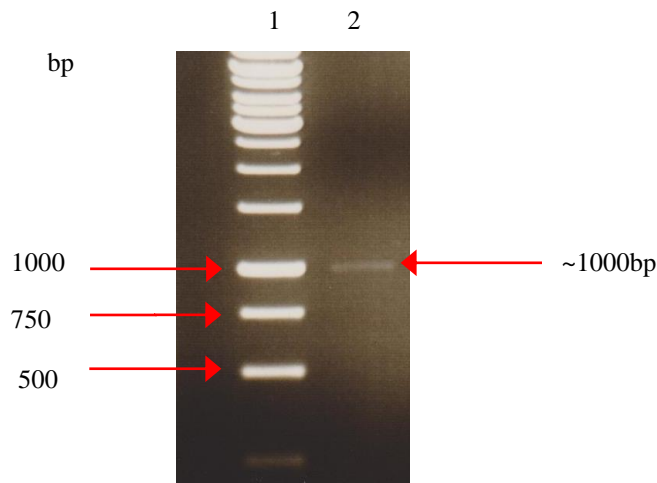


Figure 3.10 Electrophoretic Analysis of ligation reaction PCR

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2: ligation reaction PNPd PCR product with T7 forward and PNPd reverse primers

A 1% agarose gel was used to analyse the PCR products of the ligation mix through electrophoresis. In lane 2, the ~1000bp band with T7 forward and PNPd reverse

primers indicates that the PNPd PCR product is successfully ligated in pGEM vector.

Half the amount of the ligation mixture was used to transform 50µl of aliquoted XL1Blue competent bacterial cells. The cells were then plated on agar plates with the appropriate antibiotic, and were kept overnight in an incubator at 37°C. The colonies on the plates were screened by means of PCR. The PCR was done using T7 forward/ PNPd reverse primer set. PCR products were analysed by electrophoresis, and the gel is illustrated in the Figure 3.11.

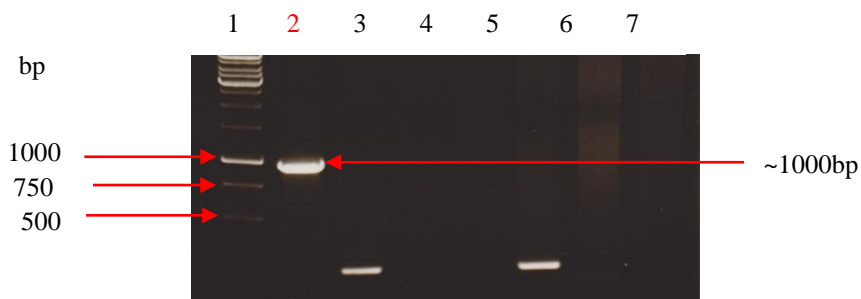


Figure 3.11 Electrophoretic Analysis of the PCR products of the pGEM vector Colony Screen

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2: Positive Colony screen PCR

Lanes 3-7: Negative Colony screen PCR

The primer set T7 forward and PNPd reverse was used for the above colony screen and the expected PCR product size is ~1000bp. Lanes 3-7 all have negative colony screen PCR products, but lane 2 has a band size of ~1000bp (~52bp for the wild type and 900bp for the PNP1 PCR product).

The colony from Figure 3.11 lane 2 was considered to be a positive colony. The whole colony was picked up from the agar plate and grown in 20ml of LB for mini prep, with the appropriate antibiotic. After mini prep the cloned plasmid was analysed by PCR, first with T7 forward/ Sp6 reverse and then with T7 forward/ PNPd primer set. The PCR products of mini preps were analysed by gel electrophoresis, and are shown in Figure 3.12. A

positive colony was confirmed by PCR and the plasmid was sent to Source Bioscience Life Sciences for sequencing.

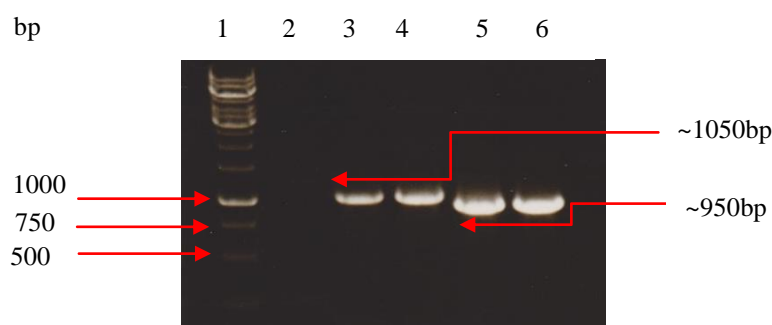


Figure 3.12 Electrophoretic Analysis of the *R. temporaria* cDNA PCR products of the cloned insert

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2: empty

Lane 3-4: Colony 2 PCR product of the pGEM vector using primer set T7 forward with SP6 reverse

Lane 5-6: colony 2 PCR product of the pGEM vector using primer set T7 forward with PNPd Reverse

To confirm that the plasmids had the right insert in the right orientation, the PCR reaction was conducted on pGEM-PNPd plasmid with T7 forward/Sp6 reverse and T7 forward/PNPd reverse. If a PNPd PCR product has been inserted in pGEM, the size of the PCR product expected with T7 forward/Sp6 reverse primer set is ~1050bp and with T7 forward/PNPd reverse primer set is ~ 950bp. The electrophoretic analysis illustrates that the resultant PCR products are of the size expected, thus confirming that the plasmids had the correct insert.

3.5.2 Sequence Analysis of the PNP1 &PNPd products for *R. temporaria*

The sequence of *R. temporaria* PNP cDNA was analysed and compared with that of *X. tropicalis* sequence found at the NCBI website, accession numbers NM_213665.1 as well as with the cloned sequence of *X. laevis* discussed in section 3.4. A sequence comparison shows that the cloned *R. temporaria* cDNA used during the project had ~99.9% similarity with the sequence found in literature for *X. tropicalis* Parapinopsin and with *X. laevis* Parapinopsin cDNA sequence. The sequence of *R. temporaria* together with *X. tropicalis* is shown below in Figure 3.13. The sequence analysis shows ~99.9% sequence similarity. The dissimilarity is due to a point mutation from G to A at a place

corresponding to 2nd position in the 156th codon. This point mutation is important due to the creation of a stop codon. To confirm this point mutation the whole experiment was repeated many times on different batches of tadpoles collected every year, different stages of tadpoles between stage 53 to small individual froglet, with different taq polymerases and with different sets of primers. Every time the sequencing results showed a point mutation at the same position, which confirms the point mutation is genuine and not due to error by taq polymerase or by the sequencing reaction. The sequencing results are shown in Table 3.1 below. (In chapter 5 Table 5.1 the complete amino acid sequence of *X. tropicalis* is shown. It can be demonstrated that the mutation in *R. temporaria* corresponds in position to 2/ 156th codon. From now onward position of this mutation will be referred to as the 2nd position of the 156th codon).

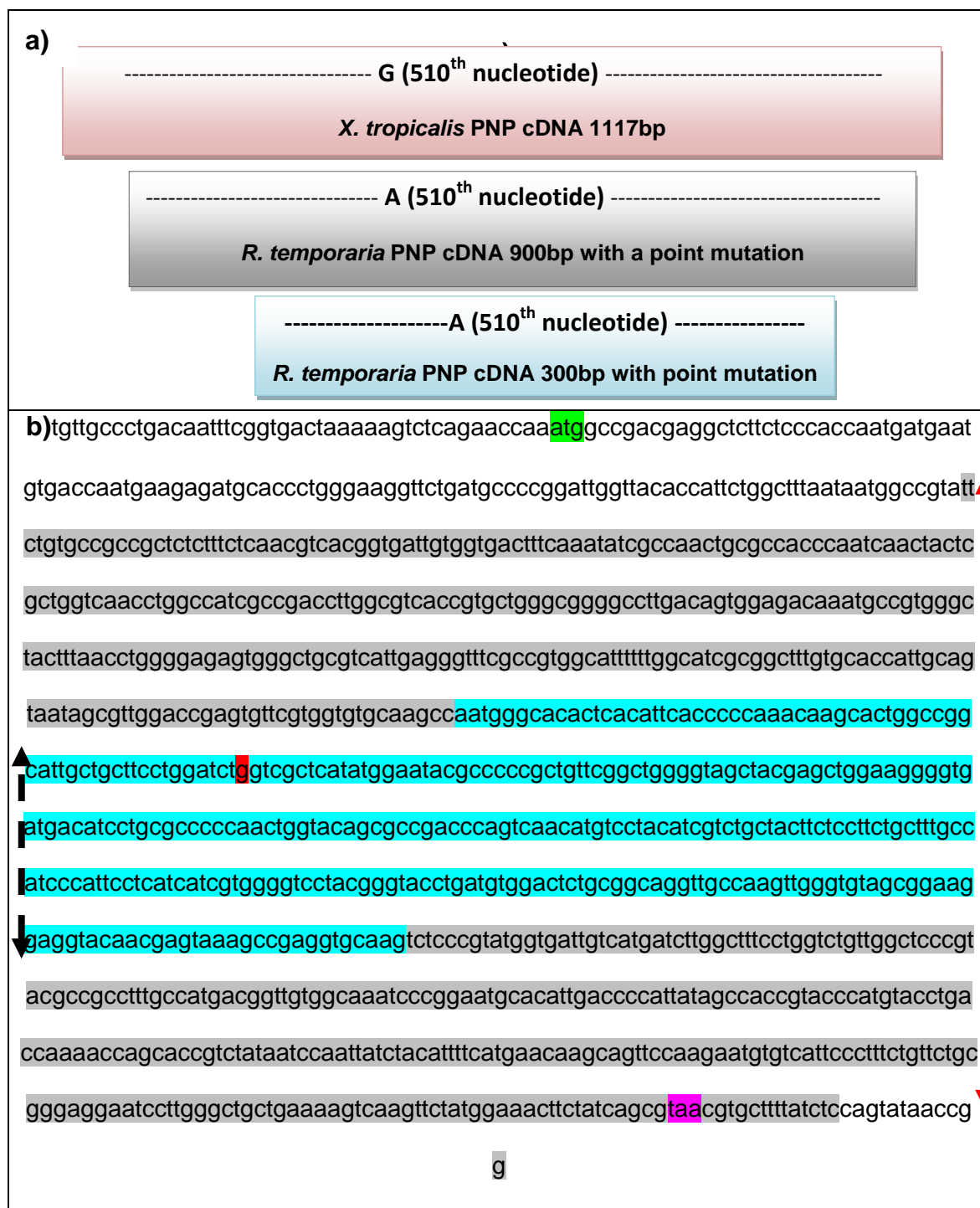
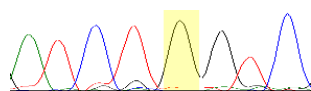
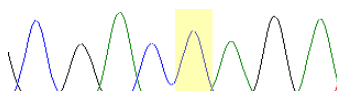
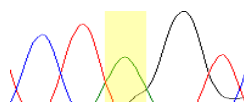
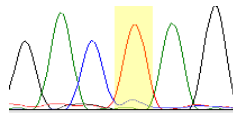
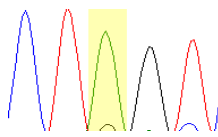


Figure 3.13 *R. temporaria* sequence matches with *X. tropicalis* Parapinopsin cDNA sequence

In the above Figure part “a” 1117bp is the PNP cDNA sequence of *X. tropicalis* reported at NCBI. *R. temporaria* cDNA 900p and 300p sequences match with *X. tropicalis* PNP cDNA except a point mutation. In part “b” the Blue highlighted region shows the *R. temporaria* 320bp sequence while grey plus blue highlighted region shows *R. temporaria* 900bp sequence; novel cDNA sequences. Both *X. tropicalis* and *R. temporaria* cDNA sequences show ~99.9% similarity. 0.1% dissimilarity is due to a point mutation at the 510th nucleotide G → A, highlighted red. Green and pink highlighted nucleotides shows start and stop codons respectively.

a) <i>X. tropicalis</i> cDNA sequence			
Name of primer		Sequence results of cDNA fragment (5'→3')	Peaks (A,C,G,T)
PNP1	Forward	ATCT G GTC	A T C T G G T C 
	Reverse	CGAC C AGA	C G A C C A G A 
b) <i>R. temporaria</i> cDNA sequence			
Name of primer	Sample details	Sequence results of cDNA fragment (5'→3')	Peaks (A,C,G,T)
1)Pnp (O)	Primer set amplified 269bp product in PCR reaction. (Batch#April,2012)	Forward CT A GT	
		Reverse GAC T AG	
2)PNPd	Primer set amplified cloned cDNA 900bp product in a PCR reaction. (Batch#April,2011)	Forward CT A GT	

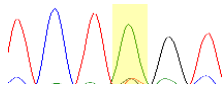
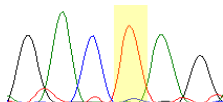
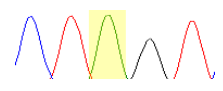
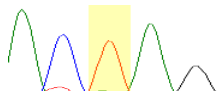
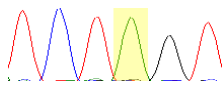
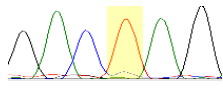
3)Pnprana	Forward	734bp PC R product (Batch#April,2010)	TCTAGT	
	Reverse		GACTAG	
4)PNP1	Forward	320bp PCR product (Batch# Apri,2009)	CTAGT	
	Reverse		ACTAG	
5) PNPexon2	Forward	181bp PCR product (Batch# Apri,2013)	TCTAGT	
	Reverse		GACTAG	

Table3.1 Sequencing results of the region around 156th codon from various cDNA samples of *R. temporaria* Parapinopsin

The Table above summarises the sequencing results of Parapinopsin cDNA. Part “a” of Table shows the sequence of *X. tropicalis* which was used as a reference sequence. Part “b” shows the sequences for *R. temporaria* Parapinopsin cDNA sequences. The first Column indicates the names of different forward and reverse primers (Primers sequence are given in chapter 2; Table 2.2). Each sample detail is given in the second column. Every time 3 tadpoles were used for mRNA extraction as per Qiagen kit instruction. cDNA sequence results are shown in the 3rd column where the red highlighted nucleotide is nucleotide of interest and indicates a point mutation in the case of *R. temporaria*. The 4th column shows the chromatogram of the sequence where the yellow highlighted peak is the peak of interest. Chromatogram peak colour codes are: A: green, C: blue, G: black and T: red. Sequencing results showed a point mutation G→A in *R. temporaria* when compared with *X. tropicalis* Parapinopsin cDNA.

3.5.3 Genotype and Quality value calculation for *R. temporaria* cDNA peak at 2nd position of the 156th codon

Table 3.2 below explains the genotype and quality value (QV) explanation. The QV is calculated by using the Applied Biosystems Sequence Scanner v 1.0 software, based on **Quality Value (QV) = $-10\log_{10} (Pe)$** where Pe is Probability of Error. Peaks with High QV represent high quality pure bases while peaks with low QV, or high probability of error, may characterize true heterozygous bases or pure bases with high background noise.

Sample details	Genotype (Sequencing result)		Quality value (QV)	
	Forward primer	Reverse primer	Forward primer	Reverse primer
<i>R. temporaria</i> (cDNA)	A	T	59	61
	A	T	61	59
	A	T	60	59

Table 3.2 Genotype and QV % for *R. temporaria* cDNA at 2nd position of 156th codon

The above Table shows the genotype and average QV assigned to 2nd position of 156th codon chromatogram for *R. temporaria* cDNA. QV is calculated by using Applied Biosystems Sequence Scanner v 1.0 software. High QV shows the high quality pure base and low QV value shows heterozygosity or high background noise.

The QV data in the above Table indicates high QV's representing high quality pure base at the 2nd position of the 156th codon. These QV were used to generate the bar chart below.

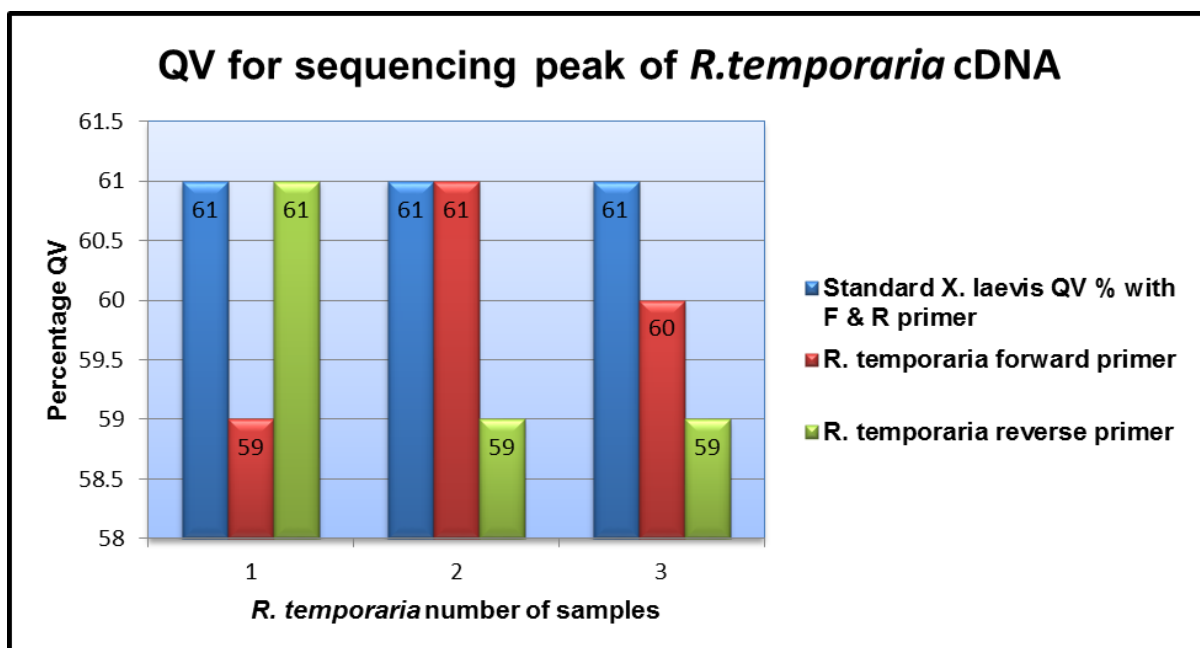


Figure3.14 The percentage of QV for sequencing *R. temporaria* cDNA at the peak of the 2nd position of 156th codon

The above figure shows the percentage QV for *R. temporaria* cDNA by taking *X. laevis* QV percentage as standard. The above figure explains the QV percentage for *R. temporaria* forward and reverse primer of samples 1, 2 and 3 referred in Table 3.2.

QV for *X. laevis*, taken as standard, was high and this QV represents the high quality pure base at 2/156th codon position. QV for *R. temporaria* cDNA is almost equal to the standard QV, representing a high quality pure base.

3.6 *Rana Tigrina* Parapinopsin cDNA study

For this part of the project, frogs were collected from Pakistan. The *R. tigrina* is a common garden frog like *R. temporaria*, but found in Pakistan, Sri-lanka, India and Bangladesh.

Total RNA was extracted from adult frog's brain. The cDNA was prepared by reverse transcription. Synthesised cDNA was then treated with DNase to remove DNA contamination. The cDNA synthesis of *R. tigrina* was confirmed for the housekeeping gene Actin with ActinXI forward and reverses primer (result not shown). A PCR reaction was conducted by using primer set PNPexon2(I) (Primers sequences are shown in Table 2.2, chapter 2). A 181bp PCR product size with PNPexon2 set would be expected. The

PCR product is shown in Figure 3.15. The PCR product using PNPExon2 sets on *R. tigrina* cDNA was then purified quantified and sent for sequencing.

3.6.1 Sequencing of *Rana tigrina* Parapinopsin cDNA

The sequence results of *R. tigrina* were analysed and compared with the *X. tropicalis* sequencing results. Sequence comparison showed ~99.9% similarity and 0.1% dissimilarity between *X. tropicalis* and *R. tigrina*. This dissimilarity is due to G→A point mutation at the 2nd position of the 156th codon. This point mutation is exactly the same as found in *R. temporaria*. Hence both *R. temporaria* and *R. tigrina* Parapinopsin sequences are identical. Sequence results of *R. tigrina* are shown in Table 3.3.

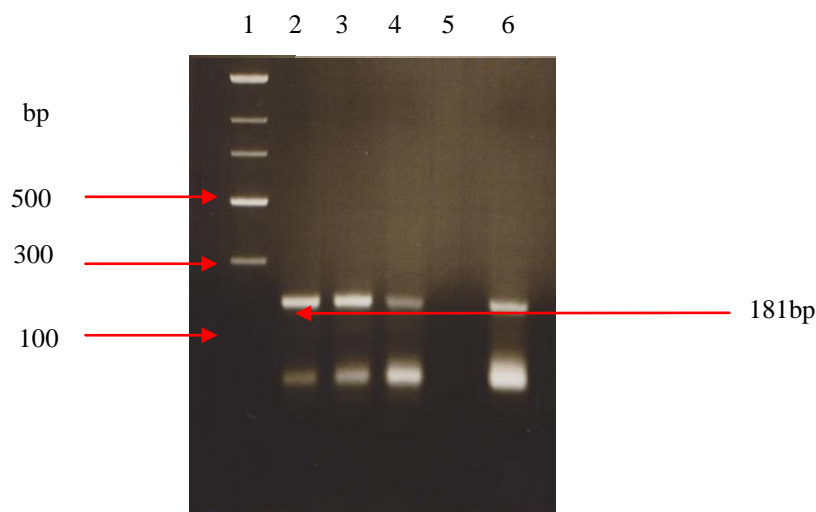


Figure 3.15 Electrophoretic Analysis of the *R. tigrina* cDNA PNPExon2 (I) PCR products

Lane 1: GeneRuler™ Express DNA ladder

Lane 2, 3, 4, 6: *R. tigrina* cDNA PCR product of PNP exon2(I)

Lane 5: empty

A 2% agarose gel was used to analyse PNP exon2(I) PCR products through electrophoresis. In lanes 2, 3, 4 and 6, the 181bp bands indicate that the cDNA of *R. tigrina* adult frog brain contains Parapinopsin. Lane 5 is empty.

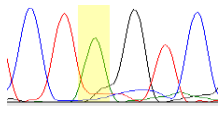
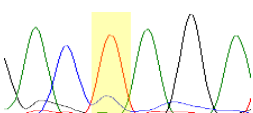
Name of primer	Sequence results of cDNA fragment (5'→3')	Peaks (A,C,G,T)
Forward	CTAGTC	
Reverse	ACTAGA	

Table3.3 Sequencing results around 156th codon of *R. tigrina* cDNA Parapinopsin

The Table shows the name of the forward and reverse primer in the 1st column, primer sequence is given in chapter 2, Table 2.2. In the 2nd column a stretch of sequence around 156th codon of Parapinopsin cDNA of *R. tigrina* is given and the mutated nucleotide is highlighted red. In column 3, a snapshot of the chromatogram is given where the yellow highlighted peak is the peak of interest. Chromatogram peak colour codes are: A: green, C: blue, G: black and T: red.

3.7 Parapinopsin Genomic DNA study

After the findings of a point mutation in *R. temporaria* and *R. tigrina* at the cDNA level, it was decided to study the Parapinopsin gene. The major problem with Parapinopsin genomic DNA study is that Parapinopsin gene sequence of *X. tropicalis* is not fully known. Less than half of the Parapinopsin cDNA sequence has been found in gDNA, only the last two introns/exons are marked in Parapinopsin gDNA reported at UCSC genome browser. The *X. tropicalis* cDNA sequence with the last two exons highlighted is shown below in Figure 3.16 b. The only complete annotated Parapinopsin gene sequence reported at UCSC genome is Zebrafish Parapinopsin (*Danio rerio* Parapinopsin b), comprising four exons; the cDNA sequence is shown in Figure 3.16 (a) below. Multiple sequence alignment (MSA-EBI) results between *X. tropicalis* / *X. laevis* and Zebrafish Parapinopsin protein show 47% homology, and more than 50% homology at the cDNA level. Hence the *X. tropicalis* cDNA sequence was divided into four exons by correlation with the intron/exon boundaries in Zebrafish Parapinopsin cDNA. According to this intron/exon division, the region of interest where the point mutation is

found in *R. temporaria* and *R. tigrina* is present in exon 2. Hence *X. tropicalis* Parapinopsin specific primers were designed very close to the mutated region within exon 2 and used for a PCR reaction.

a) Zebrafish Parapinopsin b cDNA sequence with highlighted exon boundaries

```

atg gacttttccagaattctcaacgactcttcactcactatgatggcctggttgaaggtgcctctccccgggcagg
attcatcacgctgtccctcctcatggcgtgttctccatcacatctgtcgtgtgaacgctacagtgatagtcgtcacctgcg
gcacaaacagctgcggcagccgctcaattttgccctggttaacctcgccgtggccgacctgggcaccacactaacagg
aagcgtgccatctgtggtgaccaatgctgtaggctactacattatgggacggatcggctgtgtattggagggattttgtgtg
catttttgtatctccgcgctgtgcaccgtggctctgatcgcagtgaggcgtctgtttgtggtgtgcagacctctgggctccat
cacgtttcagtgagacacgccgaggaggcctgttctcctgctggcttggctctgatctggaacactccgcctctgctg
ggctgggggaagctaccagctggagggagcgggcacgtcctgcgggccgactggcagagtcgagaactcagaga
cgtgtcttacatcatctgctacttctccgtctgcttcgggtgccattcgccatcatcctggtgtcgtattcatggctgctctaca
cgctgagacaa gtttactggaaatggtaaacaatctgggaacttttgcattgctctttcagcatactacgatttgggacat
atttcaaacactataagcattgcaacgcagcttaacgttagttttattttgtgtgaatatcactgggtgccttacgctgctt
cgctctaaccgtcgtctccaaacctgaggtccagctggccgtgctggtcaagggtttacccatatacatggccaagagca
gcactgtatacaaccccttgatctacatacatgaacaaacaggtgagcagaacaagctga

```

E1
E2
E3
E4

b) *X.tropicalis* Parapinopsin cDNA sequence with highlighted exon boundaries

```

TGTGCCCTGACAATTCGGTGACTAAAAAGTCTCAGAACCAAATCGCCGACGAGGCTCTTCTCCACCAATGATGAATGT
GACCAATGAAGAGATGCACCTGGGAAGTTCTGATGCCCCGATTGGTTACACCATTCTGGCTTTAATAATGGCCGTAT
TCTGTGCCGCCGCTCTCTTCTCAACGTCACGGTGATTGTGGTGACTTTCAAATATCGCCAACCTGCGCCACCAATCAACTA
CTCGCTGGTCAACCTGGCCATCGCCGACCTTGGCGTCACCGTGCTGGGCGGGGCCCTTGACAGTGAGACAAATGCCGTG
GGCTACTTTAACCTGGGGAGAGTGCGTGCGTCATTGAGGGTTTCGCCGTGGCATTITTTGGCATCGCGGCTTTGTGCAC
CATTGCAGTAATAGCGTTGGACCGAGTGTTCTGTTGTGCAAGCCAATGGGCACACTCACATTACCCCCAAACAAGCAC
TGGCCGGCATTGCTGCTTCTGGATCTGTCGCTCATATGGAATACGCCCGCTGTTGCGCTGGGGTAGCTACGAGCTG
GAAGGGGTGATGACATCCTGCGCCCCAACTGGTACAGCGCGACCCAGTCAACATGTCCTACATCGTCTGCTACTTCTCC
TTCTGCTTGGCATCCATTCTCATCATCGTGGGGTCTACGGGTACCTGATGTGGACTCTGCGGCAGGTTGCCAAGTTG
GGTGTAGCGGAAGGAGGTACAACGAGTAAAGCCGAGGTGCAAGTCTCCCGTATGGTGATTGTCATGATCTTGGCTTCTCT
GGTCTGTTGGTCCCGTACGCCGCTTTGCCATGACGTTGTGGCAATCCCGGAATGCACATTGACCCATTATAGCCAC
CGTACCCATGTACCTGACCAAAACCAGCACCGTCTATAATCCAATTATCTACATTTTCATGAACAAGCAGTTCCAAGAATGT
GTCATTCCCTTTCTGTTCTGCGGGAGGAATCCTTGGGCTGCTGAAAAGTCAAGTTCTATGGAACTTCTATCAGCGTAACG
AGTGGAACGCCTACGAAACGTGGCCAAGTGGCTCCGGCGTAACTGCTTTTATCTCCAGTATAACCGG

```

E3
E4

Figure 3.16 Highlighted Exons in Zebrafish and *X. tropicalis* Parapinopsin cDNA Sequence

In the above Figure part “a” shows the cDNA sequence of Zebrafish Parapinopsin b comprising 4 exons (E1, E2, E3 & E4) highlighted by alternate grey and blue colour codes. In part “b” of above Figure the last two exons 3 (E3) and exon 4 (E4) are highlighted by two colour codes, grey and blue. Rest of uppercase text within blue brackets is not known / annotated in gDNA. The last 26 nucleotide in red colour text shows the promoter region. Start and stop codons are highlighted with pink and green colour while the mutated nucleotide in *R. temporaria* / *R. tigrina* is highlighted in red. (<http://genome-euro.ucsc.edu>)

3.7.1 Exon 2 Sequencing of *Xenopus Laevis* Parapinopsin gDNA

gDNA was extracted from whole *X. laevis* tadpoles and quantified. The sets of primers PNPexon2 (O) + PNPexon2 (I) were designed based on the known sequence from *X. tropicalis* and their positions are shown in appendix 7, close to a point mutation . A PCR reaction was carried out on the gDNA (200-300ng) of *X. laevis*. A 181bp and 269bp PCR products size using PNPexon2(I) and PNPexon2(O) would be expected to be produced. PCR reaction results are shown in Figure 3.17 conducted on the gDNA of *X. laevis* on next page.

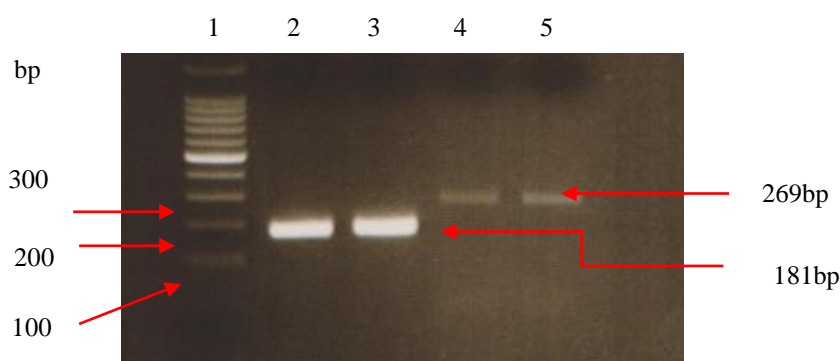


Figure 3.17 Electrophoretic Analysis of the *X. laevis* gDNA PNPexon2 PCR products

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2, 3: *X. laevis* gDNA PNPexon2(I) PCR product

Lane 4, 5: *X. laevis* gDNA PNPexon2(O) PCR product

A 2% agarose gel was used to analyse the PNPexon2(I) and PNPexon2(O) PCR products through electrophoresis. In lanes 2, 3, 4, and 5 the 181bp and 269bp bands confirm amplification of the gDNA of *X. laevis* tadpoles.

The sequencing result of *X. laevis* gDNA PNPexon2 PCR product was analysed and compared with *X. tropicalis* Parapinopsin cDNA reference sequence. The sequence

result of *X. laevis* gDNA showed 100% homology with *X. tropicalis* Parapinopsin cDNA. Sequencing results are shown in Table 3.4.

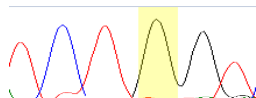
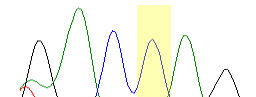
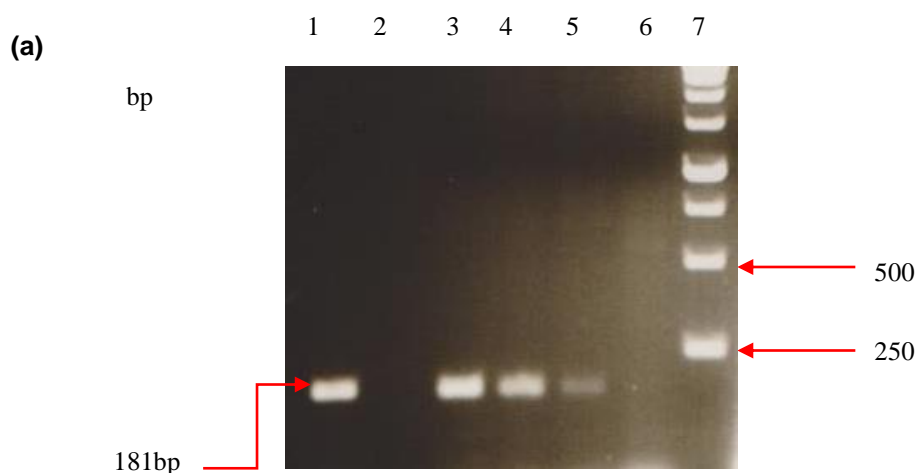
Name of primer	Stretch of Sequence results (<i>X. laevis</i> gDNA 5'→3')	Peaks (A,C,G,T)
Forward	ATCT C GT	
Reverse	CGAC C AG	

Table3.4 Sequencing results around 156th codon of gDNA of *X. laevis* Parapinopsin

The above Table shows the name of the forward and reverse primer in 1st column. The primer sequence is given in chapter 2, Table 2.2. In the 2nd column a stretch of Parapinopsin gDNA sequence of *X. laevis* is given; where the red highlighted nucleotide is the nucleotide of interest. In column 3, a snapshot of the chromatogram is given where the yellow peak is the peak of interest. Chromatogram peak colour codes are: A: green, C: blue, G: black and T: red.

3.7.2 Sequencing of *Rana temporaria* Parapinopsin gDNA

A PCR based study of *R. temporaria* gDNA was carried out by using the PNPexon2(I) and PNPexon2(O) primers (primers sequence were given in Table 2.2 Chapter 2) . 181bp and 269 bp PNPexon2(I) and PNPexon2(O) PCR products respectively would be expected. The PCR products were analyzed by gel electrophoresis, and are shown in Figure 3.18. The right size products, once confirmed, were sent for sequencing to Source Bioscience Life Sciences.



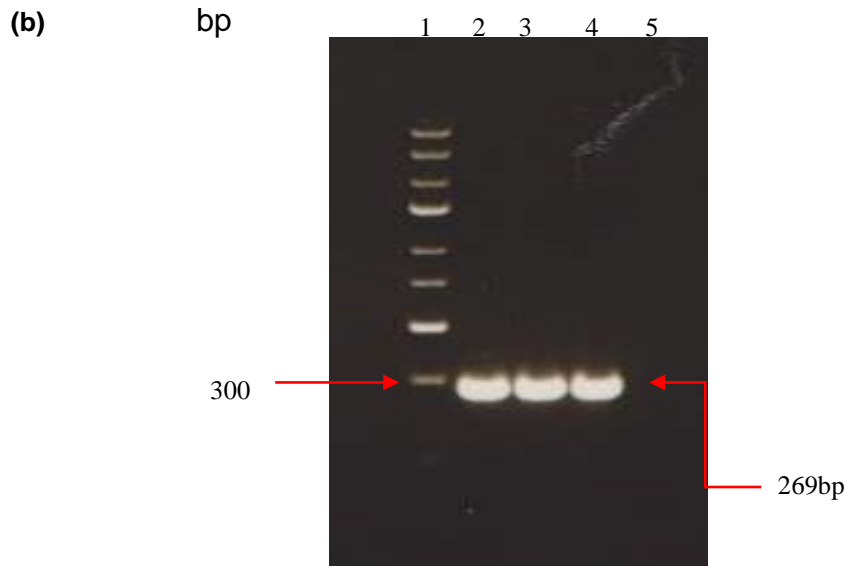


Figure 3.18 Electrophoretic Analysis of the *R. temporaria* gDNA PNPexon2(I) (a) and PNP exon2(O) PCR products (b)

(a)

Lane 1, 3, 4, 5: *R. temporaria* gDNA PCR product of PNPexon2(I)

Lane 2: PCR did not work

Lane6: Negative control; PCR without DNA template

Lane 7: GeneRulerTM 1kb DNA Ladder

A 2% agarose gel was used to analyse PNP exon2(I) PCR products through electrophoresis. In lanes 1, 3, 4 and 5 the size of 181bp bands indicates that the gDNA of *R. temporaria* has been amplified. In lane6 negative control PCR indicates that there is no amplification.

(b)

Lane 1: GeneRulerTM Express DNA ladder

Lane 2, 3, 4: *R. temporaria* gDNA PCR product of PNPexon2(O)

Lane5: Negative control

A 2% agarose gel was used to analyse PNPexon2(O) PCR products through electrophoresis. In lanes 2, 3 & 4 the size of 269bp bands indicates that the gDNA of *R. temporaria* has been amplified. In lane5 negative control PCR indicates that there is no amplification.

The sequencing results of PNPexon2(I) and PNPexon2(O) PCR products were analyzed from *R. temporaria* and compared with those from *X. tropicalis* Parapinopsin cDNA. Results are shown in Table 3.5 below. The PNPexon2(I) and PNPexon2(O) PCR was repeated many times with different lots of *R. temporaria* tadpoles' gDNA. With PNPexon2(I) reverse primer, it was observed that there is a heterozygosity at 2nd position in the 156th codon in gDNA (C, T), where a point mutation was found at cDNA level. Two times both nucleotide "C" and "T" SNP peak heights were equal (sample number 1&2). Two clear peaks of slightly different heights were observed at 2/156th in sample 3. For the rest of the time different peak heights for "C" and "T" at SNP point were obtained with both sets of primers.

Sample number	Name of primer	Stretch of Sequence results (<i>R. temporaria</i> gDNA 5'→3')	Peaks (A,C,G,T)
1	PNPexon2(I)	Forward ATCT C GTC	
		Reverse GAC C AGAT	
2	PNPexon2(I)	Forward CT C GT	
		Reverse AC C AGAT	
3	PNPexon2(I)	Forward CT A GTC	
		Reverse GAC T AGA	
4	PNPexon2(O)	Forward TCT A GT	
		Reverse GAC T AG	
5	PNPexon2(O)	Forward CT A GT	
		Reverse AC T AG	
6	PNPexon2(I)	Forward CT A GTC	

7	PNPexon2(I)	Reverse	GAC T AGAT	
		Forward	TCT A GT	
		Reverse	CGAC T AGA	
		Forward	CT A GT	
8	PNPexon2(I)	Forward	CT A GT	
		Reverse	GAC T AGAT	

Table3.5 Polymorphic sequencing results around 156th codon of Parapinopsin from 8 samples of *R. temporaria* gDNA

The Table shows the sample number and name of the forward and reverse primer in the 1st two columns. The primer's sequence is given in chapter 2, Table 2.2. Each sample contained gDNA from extraction of 3 tadpoles. In the 3rd column a stretch of Parapinopsin gDNA sequence around 156th codon is given; where red highlighted nucleotide is nucleotide of interest. In column 3, a snapshot of the chromatogram is given where yellow colour highlighted peak is the peak of interest. The sequencing results chromatogram shows SNP with the reverse primers of PNPexon2(I) and PNPexon2(O). Chromatogram peaks colour codes are: A: green, C: blue, G: black and T: red.

3.7.3 Genotype and QV calculation for *R. temporaria* gDNA at 2nd position of 156th codon

Table 3.6 below explains the genotype and quality value explanation. The 1st column shows the sample and primer detail. The QV value was calculated by using the Applied Biosystems Sequence Scanner v 1.0 software, based on **Quality Value (QV) = -10Log₁₀ (Pe)** where Pe is Probability of Error. Peaks with High QV represent high quality pure bases, while peaks with low QV, or high probability of error, may characterize true heterozygous bases or pure bases with high background noise.

<u>R. temporaria gDNA sample + Primer details</u>	<u>Genotype (Sequencing result)</u>		<u>Quality value (QV)</u>	
	Forward primer	Reverse primer	Forward primer	Reverse primer
1. PNPexon2(I)	G	C:T	18	10
2. PNPexon2(I)	G	C:T	24	11
3. PNPexon2(I)	A/G	T/C	12	24
4. PNPexon2(O)	A	T > C	35	40
5. PNPexon2(O)	A	T > C	27	24
6. PNPexon2(I)	A	T > C	35	35
7. PNPexon2(I)	A	T > C	27	24
8. PNPexon2(I)	A	T > C	24	24

Table 3.6 Genotype and QV % for *R. temporaria* gDNA at 2nd position of 156th codon

The Table shows the genotype and average QV assigned to the 2nd position of the 156th codon chromatogram for *R. temporaria*. QV is calculated by using Applied Biosystems Sequence Scanner v 1.0 software. High QV indicates the high quality pure base and low QV value indicates heterozygosity or high background noise. These QV's were used to generate the bar chart below.

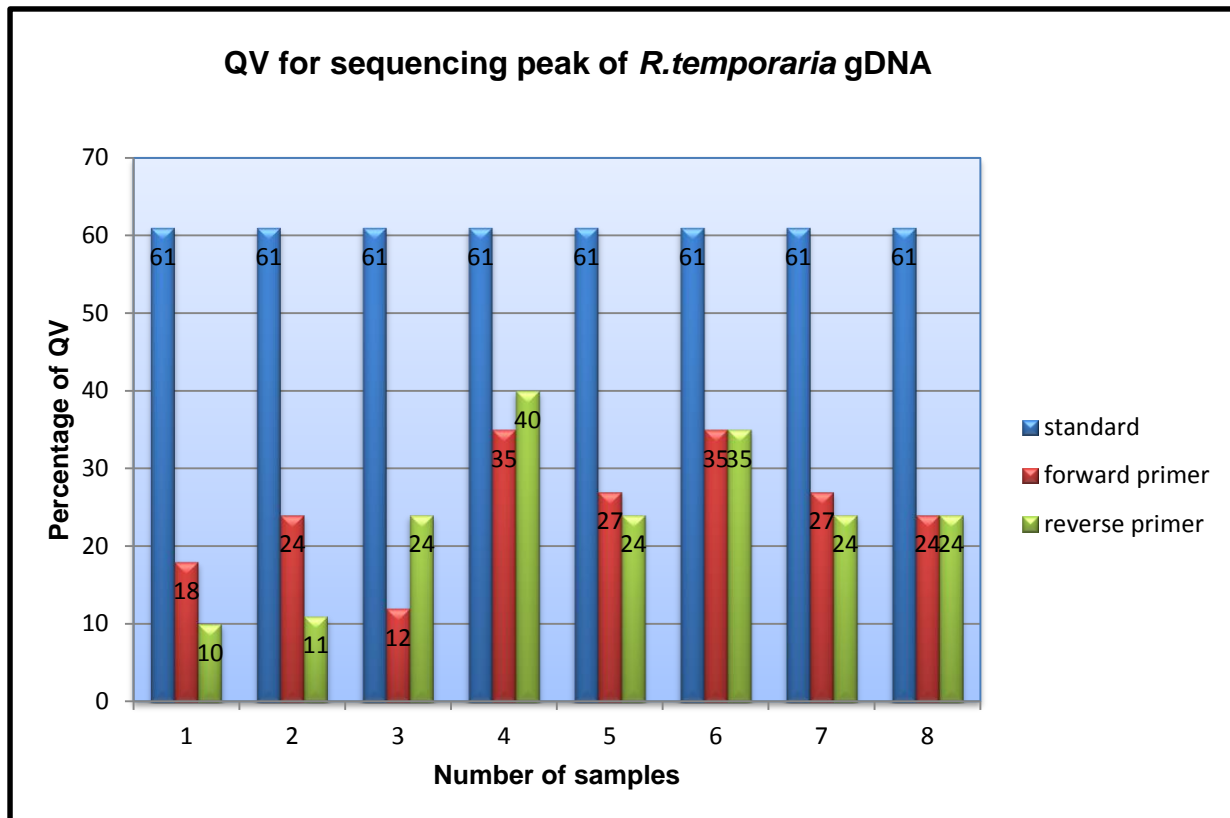


Figure3.19 The percentage QV for sequencing *R. temporaria* gDNA around the peak at the 2nd position of the 156th codon

The Figure shows the percentage of QV for *R. temporaria* gDNA by taking *X. laevis* QV percentage with forward and reverse primers as standard. Above figure shows the QV percentage at the top of each bar for forward and reverse primer of sample 1-8 inward (I) or outward (O) primers. QV for *X. laevis* was taken as standard, as this QV represents the high quality pure base at 2nd codon of 156th position. The QV for samples 1,2 and 3 were extremely low as compared to standard, (represents true heterozygosity in *R. temporaria* gDNA). Samples 4, 5, 6, 7 and 8 show higher QV as compared to samples 1, 2, and 3 but still have low QV as compared to the standard QV.

3.8 Sequencing of artificial heterozygotes (cloned *X. laevis* and *R. temporaria* cDNA mixture)

Due to unequal peak heights at the 2nd position of the 156th codon when using the primer sets, it was decided to investigate this situation further and create artificial heterozygotes with known predetermined composition. These were made by mixing cloned plasmids of *X. laevis* and *R. temporaria* Parapinopsin cDNA in exact ratios, that were then sent for sequencing to see whether the peaks would be of the same heights or

not. A plasmid with cloned *X. laevis* Parapinopsin cDNA contains “G” and a plasmid with cloned *R. temporaria* cDNA contains “A” at the 2nd position of the 156th codon. An equal mixture of these two plasmids has two copies of the gene one with “G” and the other with “A” at the said position and should give two peaks of equal height every time in the chromatogram.

Sequencing results of “artificial heterozygotes” in Table 3.7 below. Sequencing shows peaks of different heights are not always overlapping.

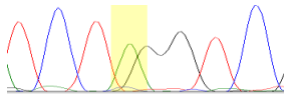
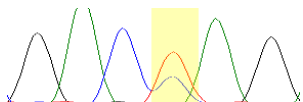
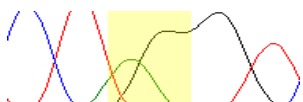
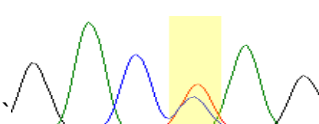
Stretch of Sequence results		
Name of primer	(cloned <i>X. laevis</i> & <i>R. temporaria</i> cDNA mixture 5'→3')	Peaks (A,C,G,T)
PNPexon2(O)	Forward ATCT T AGTC	
	Reverse CGAC T AGAT	
PNPexon2(I)	Forward ATCT NG GTC	
	Reverse CGAC T AGAT	

Table 3.7 Sequencing results around the 156th codon of 50:50 mixture of cloned cDNA from *X. laevis* and *R. temporaria*

The Table shows the name of the forward and the reverse primer in 1st column. The primer sequence is given in chapter 2, Table 2.2. In the 2nd column a stretch of the cDNA sequence of the artificial heterozygotes of cloned Parapinopsin *X. laevis* and *R. temporaria* mixture is given; where red highlighted nucleotide is nucleotide of interest. In column 3, a snapshot of chromatogram is given where yellow colour highlighted peak is the peak of interest. In the above Table two different sets of primers were used PNPexon2(I) and PNPexon2(O). Chromatogram peak colour codes are: A: green, C: blue, G: black and T: red.

3.8.1 Genotype and QV calculation for “artificial heterozygotes” at the 2nd position of 156th codon

Table 3.8 below shows the QV values from sequencing results. The QV value calculated by using the Applied Biosystems Sequence Scanner v 1.0 software, based on **Quality Value (QV) = $-10\log_{10} (Pe)$** where Pe is Probability of Error. Peaks with High QV represent high quality, pure bases while peaks with low QV, or high probability of error, may characterize true heterozygous bases or pure bases with high background noise. The QV data were used to generate a bar chart showing in Figure 3.20.

<u>Sample Details</u>	<u>Primer Name</u>	<u>Genotype</u>		<u>Quality value (QV)</u>	
		Forward primer	Reverse primer	Forward primer	Reverse primer
“Artificial heterozygotes”	PNPexon2 (O)	N:G	T/C	9:12	28
“Artificial heterozygotes”	PNPexon2 (I)	N	T/C	8	15

Table 3.8 Genotype and QV % for “artificial heterozygotes” at 2nd position of 156th codon

The Table shows the genotype, and QV assigned to the 2nd position of 156th codon chromatogram for “artificial heterozygotes”; created by mixing cloned PNP cDNA of *X. laevis* and *R. temporaria* in an equal ratio. QV is calculated by using Applied Biosystems Sequence Scanner v 1.0 software.

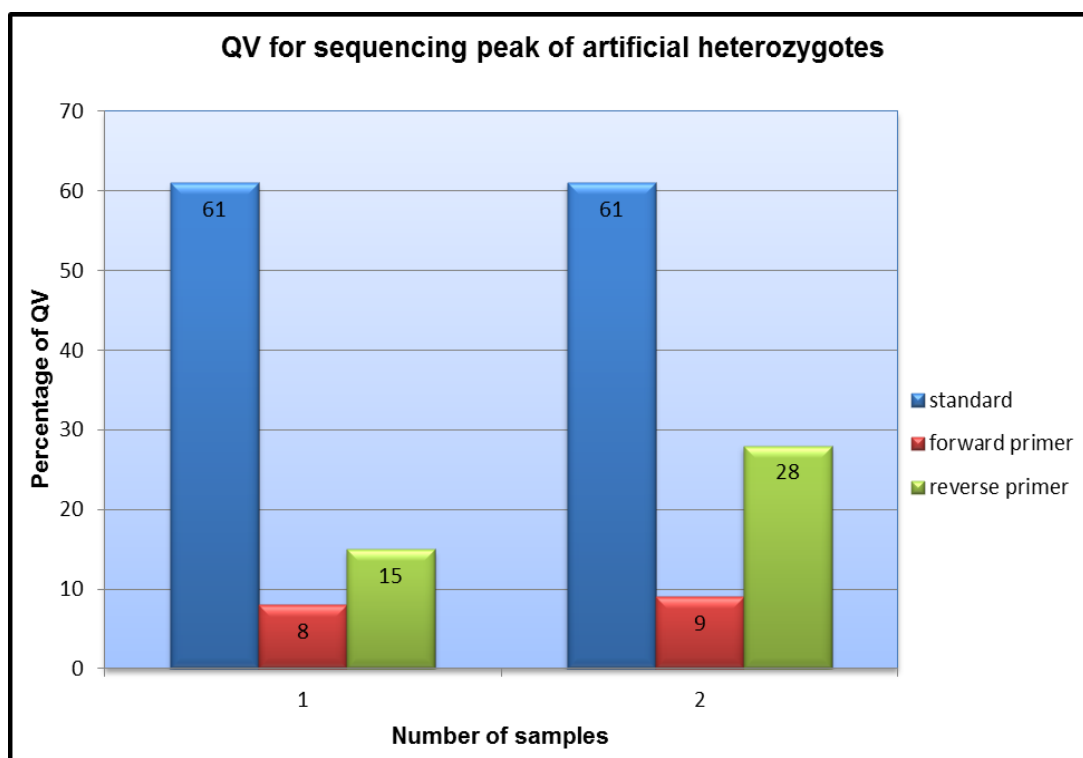


Figure3.20 The percentage QV for sequencing of “artificial heterozygotes” at 2nd position of 156th codon

The above figure shows the percentage QV for “artificial heterozygotes” by taking *X. laevis* QV percentage as standard. Above figure shows the QV percentage for forward and reverse primer for sample 1 and 2. QV for *X. laevis*, taken as standard, was high and this QV represents the high quality pure base at 2nd codon of 156th position. Lower QV for “artificial heterozygotes” as compared to standard, with forward and reverse primers represents the heterozygous situation.

To conclude, it would seem that unequal peak heights are seen upon sequencing known, exact, mixtures of cloned Parapinopsins that contain a single nucleotide heterozygous at the 2nd position of the 156th codon with the primer sets chosen. These findings indicate that the results found from sequencing gDNA from *R. temporaria* and presented in Table 3.5 and 3.6 are consistent with heterozygosity existing at this position in *Rana* Parapinopsin.

3.9 *R. temporaria* gDNA Single Tadpole Study

A PCR based study of single *R. temporaria* tadpole gDNA was carried out by using the PNPexon2(I) and PNPexon2(O) primers (primers sequence were given in Table 2.2 Chapter 2) . 181bp and 269 bp PNPexon2(I) and PNPexon2(O) PCR products respectively would be expected. The PCR products were analyzed by gel

electrophoresis, and are shown in Figure 3.21. The right size products, once confirmed, were sent for sequencing to Source Bioscience Life Sciences.



Figure 3.21 Electrophoretic analysis of the single tadpole *R. temporaria* gDNA PNPexon2(I) (a) and PNP exon2(O) PCR products

Lane 1: GeneRuler™ 100bp DNA Ladder Promega

Lane 2, 3 & 4: *R. temporaria* gDNA PCR product of PNPexon2(O)

Lane 5: Empty

Lane 6,7 & 8: *R. temporaria* gDNA PCR product of PNPexon2(I)

A 2% agarose gel was used to analyse PNP exon2(I) and PNP exon2(O) PCR products through electrophoresis. In lanes 2, 3, 4 the bands of 269bp indicate that the gDNA of *R. temporaria* contains Parapinopsin. Lane 5 is empty. In lanes 6, 7 & 8, the band of 181 bp indicates PNP exon2(I) primers amplified.

Sample number	Name of primer	Stretch of Sequence results (<i>R. temporaria</i> gDNA 5'→3')	Peaks (A,C,G,T)
1	PNPexon2(I)	Forward TCT A GT	
		Reverse GACT T AG	
2	PNPexon2(I)	Forward TCT A GT	
		Reverse GACT T AG	
3	PNPexon2(I)	Forward TCT A GT	
		Reverse GACT T AG	
4	PNPexon2(O)	Reverse GACT T AG	
5	PNPexon2(O)	Forward TCT A GT	
		Reverse GACT T AG	

Table3.9 Polymorphic sequencing results around 156th codon of Parapinopsin gDNA from 5 samples of single tadpole *R. temporaria* gDNA

The Table shows the sample number and name of the forward and reverse primer in 1st two columns. The primer's sequence is given in chapter 2, Table 2.2. Each sample contained gDNA from extraction of a single tadpole. In the 3rd column a stretch of Parapinopsin gDNA sequence around 156th codon is given; where red

highlighted nucleotide is nucleotide of interest. In column 3, a snapshot of the chromatogram is given where yellow colour highlighted peak is the peak of interest. The sequencing results chromatogram shows the small peak for wild type allele inside a high peak for mutant allele, with the reverse primers of PNPexon2(I) and PNPexon2(O). Chromatogram peaks colour codes are: A: green, C: blue, G: black and T: red.

3.9.1 Genotype and QV Calculation for *R. temporaria* gDNA at 2nd Position of 156th Codon

Table 3.10 below explains the genotype and quality value explanation. The 1st column shows the sample and primer detail. The QV value was calculated by using the Applied Biosystems Sequence Scanner v 1.0 software, based on **Quality Value (QV) = -10Log₁₀ (Pe)** where Pe is Probability of Error. Peaks with High QV represent high quality pure bases, while peaks with low QV, or high probability of error, may characterize true heterozygous bases or pure bases with high background noise.

<u><i>R. temporaria</i> gDNA sample + Primer details</u>	<u>Genotype (Sequencing result)</u>		<u>Quality value (QV)</u>	
	Forward primer	Reverse primer	Forward primer	Reverse primer
1. PNPexon2(I)	A	T	27	49
2. PNPexon2(I)	A	T	24	43
3. PNPexon2(I)	A	T	34	60
4. PNPexon2(O)	-	T	-	47
5. PNPexon2(O)	A	T	40	47

Table 3.10 Genotype and QV % for *R. temporaria* single tadpole gDNA at 2nd position of 156th codon

The Table shows the genotype and average QV assigned to the 2nd position of the 156th codon chromatogram for *R. temporaria* single tadpole study. QV is calculated by using Applied Biosystems Sequence Scanner v 1.0 software. High QV indicates the high quality pure base and low QV value indicates heterozygosity or high background noise.

The QV values from Table 3.10 were used to generate a bar chart is shown in Figure 2.22.

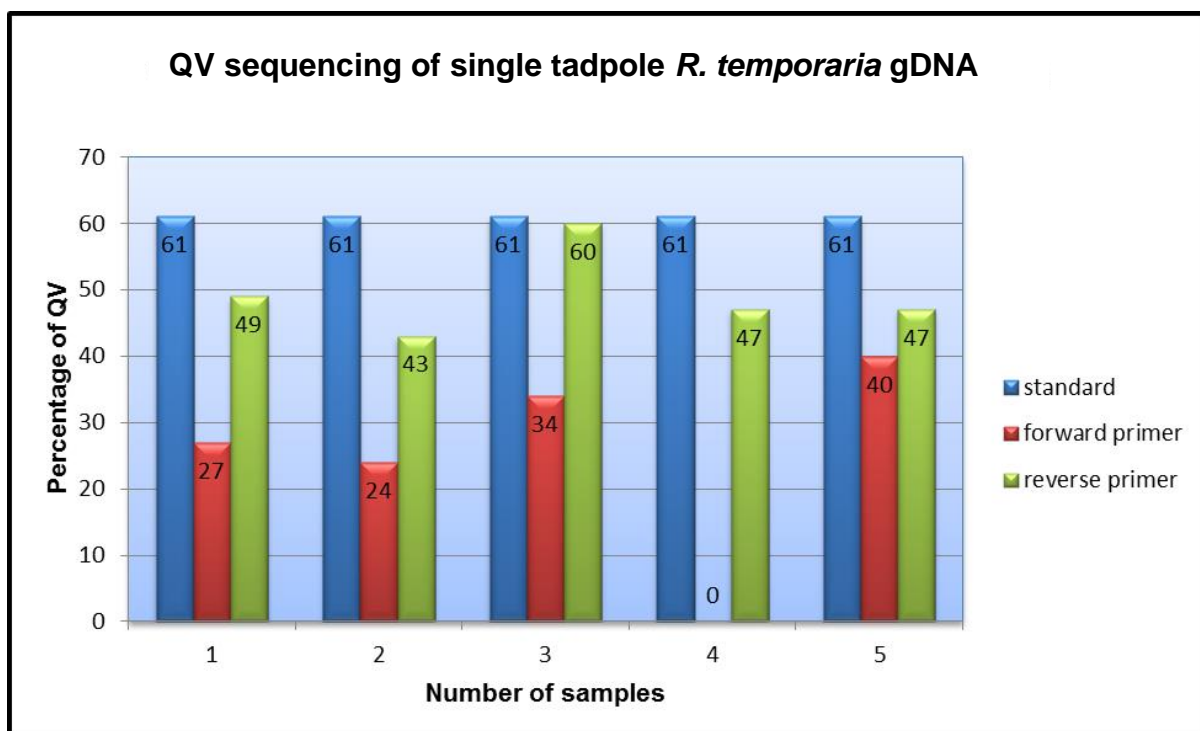


Figure3.22 The percentage QV for sequencing of single tadpole *R. temporaria* gDNA at the 2nd position of the 156th codon for Parapinopsin

The above figure shows the percentage of QV for *R. temporaria* gDNA by taking *X. laevis* QV percentage as standard. Above figure shows the QV percentage for forward and reverse primer of samples 1, 2, 3, 4 and 5. The QV's with the forward primers are lower as compared to standard but still higher than QV for "artificial heterozygotes" shown in Table 3.21. QV for *R. temporaria* gDNA with reverse primers.

3.10 Sequencing of *Rana tigrina* Parapinopsin gDNA

A PCR based study of *Rana tigrina* gDNA was carried out by using the same PNPexon2(I) and PNPexon2(O) primers that were used for the *R. temporaria* gDNA study. 181bp and 269 bp size PNPexon2(I) and PNPexon2(O) PCR products would be expected. PCR products were analysed by gel electrophoresis and are shown in Figure 3. 23. The right size product, once confirmed, was sent for sequencing to Source Bioscience Life Sciences. Results are shown in the Table 3.9.

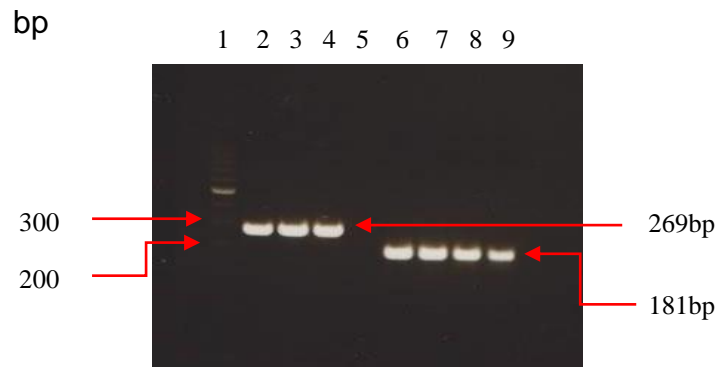


Figure 3.23 Electrophoretic Analysis of *R. tigrina* gDNA PNPexon2(I) & PNPexon2(O) PCR products

Lane 1: 100bp DNA Ladder, Promega

Lane 2, 3 & 4: *R. tigrina* gDNA PCR product of PNPexon2(O)

Lane5: Negative control PCR without template

Lane 6, 7, 8 & 9: *R. tigrina* gDNA PCR product of PNPexon2(I)

A 2% agarose gel was used to analyse PNP exon2(I) and PNP exon2(O) PCR products through electrophoresis. In lanes 2, 3, 4 the bands of 269bp indicate that the gDNA of *R. temporaria* contains Parapinopsin. In lane 5 negative control a PCR without template showed no amplification. In lanes 6, 7, 8 & 9, the band of 181 bp indicates PNP exon2(I) primers amplified.

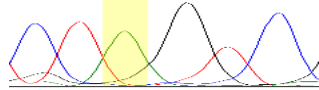
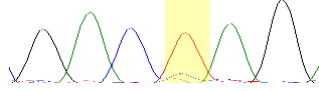
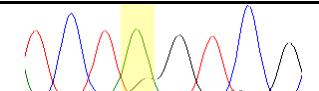
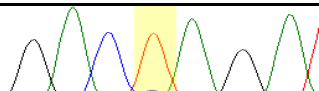
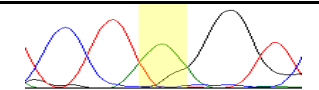
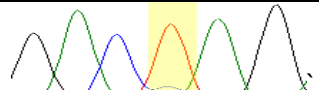
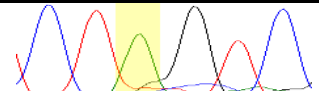
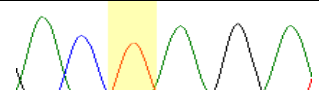
Name of primer		Stretch of Sequence results (<i>R. tigrina</i> gDNA 5'→3')	Peaks (A,C,G,T)
PNPexon2(I)	Forward	CTAGTC	
	Reverse	GACTAG	
PNPexon2(I)	Forward	TCTAGTCG	
	Reverse	GACTAGA	
PNPexon2(O)	Forward	CTAGT	
	Reverse	GACTAG	
PNPexon2(O)	Forward	CTAGTC	
	Reverse	ACTAGA	

Table3.9 Sequencing results of gDNA study around 156th codon of *R. tigrina* Parapinopsin

The Table shows the name of the forward and reverse primer in the 1st column. (The primer sequence is given in chapter 2, Table 2.2). In the 2nd column a stretch of Parapinopsin gDNA sequence of *R. temporaria* is given; where red highlighted nucleotide is nucleotide of interest. In column 3, a snapshot of the chromatogram is given where the yellow colour highlighted peak is the peak of interest. The sequencing result chromatogram shows the small peak for wild type allele inside a high peak for mutant allele, with the reverse primers of PNPexon2(I) and PNPexon2(O). In the above Table two different sets of primers were used PNPexon2(I) and PNPexon2(O). Chromatogram peak colour codes are: A: green, C: blue, G: black and T: red.

3.10.1 Genotype and QV calculation for the 2nd position of 156th codon in Parapinopsin from *R. tigrina* gDNA

The QV were calculated as previously by using **Quality Value (QV) = $-10\log_{10} (Pe)$** where Pe is Probability of Error and are shown in Table 3.10 and Figure 3.24.

<u>Sample detail</u>	<u>Primer Name</u>	<u>Genotype (Sequencing result)</u>		<u>Quality value (QV)</u>	
		Forward primer	Reverse primer	Forward primer	Reverse primer
<i>R. tigrina</i> (gDNA)	PNPexon2(I)	A	T	44	51
	PNPexon2(I)	A	T	44	49
	PNPexon2(O)	A	T	45	54
	PNPexon2(O)	A	T	46	59

Table 3.10 Genotype and QV % for at 2nd position of 156th codon of Parapinopsin from *R. tigrina* using gDNA

The Table shows the genotype and average QV assigned at the 2nd position of the 156th codon chromatogram for *R. tigrina* gDNA. QV is calculated by using Applied Biosystems Sequence Scanner v 1.0 software. High QV represents high quality pure base and low QV value represents heterozygosity or high background noise. The QV data in the above Table with high QV shows high quality pure base at 2nd position of 156th codon. These QV were used to generate the bar chart below.

The QV data here indicates high QV *i.e.* high quality pure bases at the 2nd position of the 156th codon. Strong evidence for heterozygosity, similar to that seen for *R. temporaria* gDNA samples is not apparent.

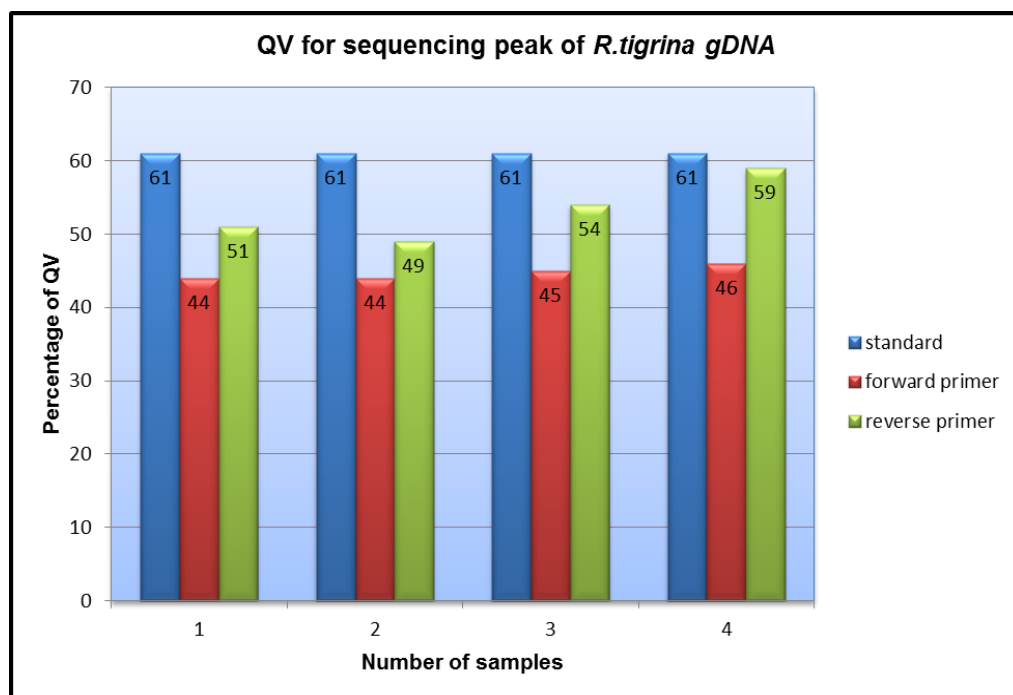


Figure3.24 The percentage QV for sequencing of *R. tigrina* gDNA at 2nd position of 156th codon for Parapinopsin

The figure shows the percentage of QV for *R. tigrina* gDNA by taking *X. laevis* QV percentage as standard. The above figure shows the QV percentage for forward and reverse primer of samples 1, 2, 3 and 4. The QV with forward primers are lower compared to standard, but still higher than QV for “artificial heterozygotes” shown in Table 3.21. QV for *R. tigrina* gDNA with reverse primers are higher and close to standard QVs which represents high quality pure base.

3.11 Summary of cDNA and gDNA Parapinopsin study Results

<u>Frog's Name</u>	<u>Parapinopsin cDNA study</u>	<u>Parapinopsin gDNA study</u>
<i>X. tropicalis</i>	Reference sequence reported at NCBI	Reference sequence reported at UCSC
<i>X. laevis</i>	Sequence results show 100% homology with <i>X. tropicalis</i> .	Sequence results show 100% homology with <i>X. tropicalis</i> .
<i>R. temporaria</i>	Sequence study shows ~99.9% similarity with <i>X. tropicalis</i> and a point mutation at 2 nd position of 156 th codon which is G→A.	Sequence study shows ~99.9% similarity with <i>X. tropicalis</i> and SNP at 2 nd position of 156 th codon which is C→T on reverse strand.
<i>R. tigrina</i>	Sequence study shows ~99.9% similarity with <i>X. tropicalis</i> and a point mutation at 2 nd position of 156 th codon which is G→A.	Sequence study shows ~99.9% similarity with <i>X. tropicalis</i> and a point mutation at 2 nd position of 156 th codon which is G→A.

Table 3.11 Summary of Parapinopsin study in frogs

Table above summarizes the results from cDNA and gDNA Parapinopsin studies.

3.12 Summary of Genotype at 2nd position of 156th codon

<u>Sample details</u>	<u>Genotype (Sequencing result)</u>		<u>Quality value (QV)</u>	
	Forward primer	Reverse primer	Forward primer	Reverse primer
<i>X. tropicalis</i> (gDNA)	G	C	59	61
<i>X. laevis</i> gDNA	G	C	61	61
<i>X. laevis</i> cDNA	G	C	61	61
<i>R. temporaria</i> (cDNA)	A	T	59	61
	A	T	61	59
	A	T	60	59
<i>R. temporaria</i> (gDNA)	Sample1 G	C:T	18	10
	Sample2 G	C:T	24	11
	Sample3 A/G	T/C	12	24
	Sample4 A	T > C	35	40
	Sample5 A	T > C	27	24
	Sample6 A	T > C	35	35
	Sample7 A	T > C	27	24
	Sample8 A	T > C	24	24
	Sample9 A	T > C	27	49
	Sample10 A	T > C	24	43
	Sample11 A	T > C	34	60
	Sample12 A	T > C	-	47
	Sample13 A	T > C	40	47
<i>R. tigrina</i> (cDNA)	Sample1 A	T	47	47
<i>R. tigrina</i> (gDNA)	Sample1 A	T	44	51
	Sample2 A	T	44	49
	Sample3 A	T	45	54
	Sample4 A	T	46	59

Artificial heterozygotes	Sample1 N:G	T/C	9	15
<i>X. laevis</i> + <i>R. temporaria</i> cDNA	Sample2 N	T/C	8	28

Table3.12 : Genotype and Quality value calculation of peaks at 2nd position of 156th codon

The Table above explains the genotype and quality value explanation. In the 1st column the names of different species of frogs are given. 2nd column explains the sequencing result at 2nd position of 156th codon with forward and reverse primer. 3rd column explains the quality values (QV) of peaks at the position with forward and reverse primer. The QV were calculated by using the Applied Biosystems Sequence Scanner v 1.0 software, based on **Quality Value (QV) = -10Log₁₀ (Pe)** where Pe is Probability of Error represent the high quality pure base while peaks with low QV or high probability of error may characterize true heterozygous bases or pure bases with high background noise. On the average, peak results of *X. tropicalis* (gDNA), *X. laevis* (gDNA, cDNA), *R. temporaria* (cDNA) and *R. tigrina* (gDNA, cDNA) gave high QV and smaller peak inside a high mutant allele/peak like *R. temporaria* single tadpole study results (sample #9-13), while all peak results of *R. temporaria* gDNA give low QV which may represent heterozygosity in *R. temporaria* gDNA. QV of samples 1, 2 and 3 of *R. temporaria* gDNA are extremely low which may represent 100% heterozygosity and two peaks at a single position. The sequence peaks are clean and there is no chance of error due to background noise. QV values are a little higher in the case of Sample 4 of *R. temporaria* gDNA as compared to the rest of 5, 6, 7 and 8 samples but still less than the rest of QV given in the Table. Samples 4,5,6,7 and 8 showed two peaks / heterozygosity but peaks of different heights.

3.13 Discussion

This chapter describes the study of the coding region of Parapinopsin from amphibian cDNA and gDNA. Parapinopsin belongs to a group of membrane proteins called G protein coupled receptors (Wada *et al.*, 2012). Parapinopsin is a non-visual photosensitive protein expressed in the pineal and parapineal part of the brain in lower vertebrates and was studied for the first time by Blackshaw and Snyder, 1997. The amphibian has remained an excellent animal for research since 1905 in developmental and cell biology (Brown, 2005). Briggs and King (1952) successfully transplanted for the first time blastula cell nuclei from *R. catesbeiana* (American Green Frog) into *R. pipiens* (Northern leopard frogs) eggs of pre- damaged nucleus. John Gurdon repeated the same experiment in 1960 on *X. laevis* and microinjected the intestinal nuclei into eggs whose nuclei had been inactivated. John Gurdon's work on *Xenopus* spanned more than 50 years and recently he revealed the fact that somatic cells can revert to stem cells (Gurdon, 2013).

The current study was conducted on four different frog species *X. tropicalis* (Western clawed frog), *X. laevis* (African clawed frog), *R. temporaria* (common European frog) and *R. tigrina* (Asian bullfrog) using local sources and by contacting different research groups who were already working on these species. For cDNA studies, tadpoles were cultured and grown to a particular stage i.e. stage 53 by using Nieuwkoop and Faber *X. laevis* tadpole development staging system established in 1967 (McDiarmid, 1999). The Parapinopsin protein is expressed in the pineal gland in brain, and brain development is complete at stage 53 (<http://www.xenbase.org/anatomy>). The gDNA and total mRNA was extracted from whole tadpoles at stage 53. The tadpoles were also preserved in liquid nitrogen at -80°C (Stanley, 2008) or in ethanol (McDiarmid, 1999) for future use. The mRNA was used as a template by reverse transcriptase to make a cDNA. The only reported cDNA sequence of Parapinopsin at NCBI is that of *X. tropicalis* and its whole pineal mRNA extract was used by Koyanagi *et al* in 2004. Hence gene specific primers were designed against *X. tropicalis* Parapinopsin cDNA and used for PCR against cDNA and gDNA from different frog's species. The PCR amplified products of *X. laevis* and

R.temporaria cDNA were cloned in pGEM vectors whereas the other PCR products from species were directly used for sequencing.

The 900bp of *X.laevis* from 163bp - 1062bp cDNA (Fig 3.7) sequence showed 100% homology with the *X.tropicalis* reported Parapinopsin sequence (Appendix 1). The flanking regions in *X. laevis* Parapinopsin from 1bp-162bp & 1062bp-1117bp were not sequenced despite various attempts to get a successful PCR using different primers sets directed against these flanking regions.. Neither were sequencing data obtained for the corresponding 5' sequence (1-162) for the *Rana* parapinopsin genes. The 100% homology novel finding is supported by Hirsch *et al*, 2002; due to their close relationship most of the sequences are conserved in *X.tropicalis* and *X.laevis*. *X.laevis* probes were used on *X.tropicalis* during developmental studies and *X.laevis* antibodies react against *X.tropicalis* protein (Burki 1985, de Sá RO & Hillis 1990 and Hirsch *et.al.*, 2002). Another novel finding during the current project was that *R.temporaria* and *R.tigrina* Parapinopsin cDNA sequences showed ~99.9% sequence similarity with *X.tropicalis*. The 0.1% dissimilarity is due to a point mutation at the 2nd position of the 156th codon which is G→A. This point mutation is very important because it creates a stop codon UGG→ UAG and hence protein truncation.

Parapinopsin is an α helical seven pass transmembrane protein (Kawano-Yamashita *et al.*, 2011) and consists of 348 amino acids (<http://www.ncbi.nlm.nih.gov/nuccore>) with a molecular weight of ~38kDa (Wada S *et al.*, 2012). Parapinopsin protein was studied in comparison with Bovine Rhodopsin protein (Akihisa Terakita, 2005). This sequence comparison information was used to find out the exact position of a point mutation in the 7 helical protein structure. It was found that a point mutation in *R.temporaria* and *R.tigrina* Parapinopsin creates a stop codon right after the 155th amino acid, located in the 4th helix.

Due to the importance of the effect of this point mutation, the whole experiment was repeated many times using different approaches to confirm that the point mutation was real: on different batches of tadpoles which were collected every year around February to

April from 2010- 2013. Each batch sample was further sub-divided and labelled into group of three tadpoles. Different primer sets downstream and upstream of the mutated region were used. Every time the results confirmed that the point mutation was real, it's not due to taq polymerase mismatch nucleotide incorporation during PCR or due to sequencing error (Simsek *et al.*, 2001) because of reproducibility. To get good, reliable and clear sequencing peaks without background noise, post PCR clean-up treatment of a sample would be required to remove extra PCR primers and dNTPs (Pui-Yan Kwok and Chen, 2003). Hence all the PCR samples were treated to remove excess of primers and dNTPs by extraction of required PCR products from gel prior to sequencing. The current studies show only high quality sequencing results with no background noise.

During the study all steps recommended for successful sequencing were carefully adhered to *e.g.*, purity of DNA samples (A260/A280 ratios 1.8 for DNA and 2 for RNA), ratios smaller than this may indicate contamination (Sambrook *et al.*, 1989), template and primer concentrations used in sequencing reaction (too low primer concentration results in low sequencing signal and too high primer concentration results in over-amplification of 5' end).

Primer sequence, length, T_m and purification of primers are also critical for successful sequencing. During the current study all primers used, were HPLC purified (Hanke and Wink, 1994, Stella Lai *et al.*, 2012).

The Parapinopsin study is at an early stage and little information is available about the Parapinopsin gene for example the position of a gene on the chromosome or the gene coding regions. Only the last two exons are marked (annotated) in gDNA of *X.tropicalis* Parapinopsin reported at UCSC genome browser (<http://genome.ucsc.edu>). A point mutation at 2nd position of 156th codon in Parapinopsin cDNA is present somewhere upstream in the gDNA which is not sequenced yet. The current study was also conducted to get more knowledge about the *X.tropicalis* Parapinopsin gene around the region mutated in *R.temporaria*, and it was concluded that the point mutation is present presumably in exon 2.

Designing PCR primers against *X.tropicalis* Parapinopsin gene around the 156th codon was one major problem due to lack of information on intron/exon boundaries, hence I started looking into the fully annotated Parapinopsin gene sequence and found Zebrafish (*Danio rerio*) Parapinopsin b gene located on chromosome 11 comprising 4 exons (<http://genome.ucsc.edu>). *X.tropicalis* cDNA was divided into four exons by lining up the sequence with Zebrafish Parapinopsin cDNA sequence. Multiple sequence alignment (MSA-EBI) results between *X.tropicalis*/*X.laevis* and Zebrafish Parapinopsin protein showed 47% similarity and more than 50% similarity at the cDNA level. Hence *X.tropicalis* cDNA sequence was divided into four exons by correlating the intron/exon boundaries with Zebrafish Parapinopsin cDNA. According to this intron/exon division, the region of interest where the point mutation is found in *R.temporaria* and *R.tigrina* is present in exon 2. Hence *X.tropicalis* and Zebrafish Parapinopsin (Zfishpnp) specific primers were designed very close to the mutated region within exon 2 and used for PCR reaction. PCR reactions were carried out on Zebrafish and *X.tropicalis*, *X.laevis*, and *R.temporaria* and *R.tigrina* gDNA. The sequence results of Zfishpnp PCR product matched those of the reported Zebrafish cDNA sequence at NCBI. The *X.laevis*, *X.tropicalis* and *R.tigrina* Parapinopsin gDNA coding sequence study results support the cDNA sequence study results from the species. But gDNA sequence study results of *R.temporaria* were different from its cDNA sequence results. gDNA sequence results from *R.temporaria* showed single nucleotide polymorphism A/G at the same position where a point mutation was observed at the cDNA level (i.e. 2nd position of 156th codon). heterozygosity in Parapinopsin of *R.temporaria* is a novel finding. To confirm heterozygosity the whole experiment was repeated many times.

Heterozygosity is very common in *R.temporaria* and important in many aspects. The heterozygote individuals are more resistant than homozygous ones, and this depends upon the size of tadpoles during metamorphosis. The bigger the size of a tadpole the greater will be the chances of survival in the natural environment and also a heterozygote female has more chances to nourish eggs with yolk than less heterozygotes female (Lesbarre`res *et al.*, 2007). Genetic variability and age

relationship is also very important in a frog's life; heterozygous individuals show high metabolic rates and have ability to survive in stressful environments (Schmeller *et al.*, 2007). The *R.temporaria* tadpoles used during the current study were collected from a garden pond and grew under natural environmental conditions. In the case of amphibians, particularly in *R.temporaria*, environmental stress has a great influence on gene expression and maternal effects in terms of viability (Pakkasmaa *et al.*, 2003).

R.temporaria sequencing results were carefully analysed and peaks heights were assigned Quality Values (QV). These QV values (Curtis *et al.*, 2010) were calculated by user friendly software the Applied Biosystems Sequence Scanner v 1.0 (Srivatsan *et al.*, 2008, Richter *et al.*, 2013, Hong Liu *et al.*, 2013), based on Quality Value (QV) = $-10\log_{10} (Pe)$ where Pe is Probability of Error (Brockman *et al.*, 2008). The software Sequencher 4.0 (Gene Codes Corporation Ann Arbor, MI) was used for SNP detection by K. Ukoskit in 2001; the software calculates each base peak height to interpret data. The Applied Biosystems Sequence Scanner v 1.0 is updated software, and peak heights were assigned QV. One can find heterozygosity by two parameters peak heights and QV by using current software. Clear single sequencing peaks with high QV represent high quality, pure bases, while double peaks with low QV, or high probability of error, may characterize true heterozygous bases, or pure bases with high background noise. QV were assigned to all sequencing results at the 2/156th codon. *X.tropicalis* (gDNA), *X.laevis* (gDNA, cDNA), *R.tigrina* (gDNA, cDNA) and *R.temporaria* (cDNA), sequencing results that give high QV and clear single peaks, with no background noise, represent high quality pure bases. The QV for peaks of *R.temporaria* gDNA were low which represents heterozygosity in *R.temporaria* gDNA. QV for samples 1, 2 and 3 of *R.temporaria* gDNA (Table 3.6) were extremely low which may represent 100% heterozygosity. Two clear peaks are seen at a single position. The sequence peaks were clean and there is no chance of the probability of error due to background noise. QVs are little higher in the case of Sample 4 of *R.temporaria* gDNA compared to the rest of samples 5, 6, 7 and 8. Samples 4,5,6,7 and 8 sequence results showed two peaks of unequal heights. Sequencing can be problematic because of dissimilar rates of

incorporation of the dideoxynucleotides which result in differences in peak height (Ukoskit, 2001, Simsek *et al.*, 2001).

Artificial heterozygotes were produced in current study by mixing cloned cDNA from *X.laevis* and *R.temporaria* in exactly the same ratio to investigate heterozygous peak heights. The sequencing results showed heterozygous peaks of different heights, but two clear peaks were seen at the same position with low QV. One reason for differences in peak heights may be due to the labelled dideoxy dye attached to C “ddC” being degraded resulting in smaller “C” peaks inside peak “A” (Table 3.7) with reverse primer (AB Sanger Sequencing Guide). The other reason is may be due to weaker hybridization of PNPexon2(O) and PNPexon2 (I) reverse primers with allele “C” resulting in smaller peak for “C” at 2/156th position (Lagerkvist *et al.*, 1994).

In *R.temporaria* gDNA samples (in table 3.5, samples number 4,5,6,7 and 8) peak height differences are high as compared to results from artificial heterozygote table 3.7 and this supports the already reported literature review by Matsuba and Merila, 2006 about genetic variability present in *R.temporaria*. The current study may bridge the occurrence of loss of heterozygosity with reported genetic variability in *R.temporaria*. Loss of allele “C” peak height inside allele “T” peak was observed in samples 4,5,6,7 and 8. Loss of heterozygosity was observed through sequencing in ovarian melanomas, breast carcinomas and colorectal carcinomas (Musani *et al.*, 2012, Murthy *et al.*, 2002, Zavodna *et al.*, 2009).

Sanger sequencing is the “Gold Standard” for mutation detection/heterozygosity. However it detects mutations at a certain allelic level. Studies reveal that a lower level of mutation, due to higher ratio of normal to mutant DNA, is not detected by sequencing. Mutations are only detected by sequencing at certain sequencing sensitivity levels, which are 10-20% mutant to normal (Hongdo and Dobrovic, 2009).

The results from the current study in the light of reviewed literature are significant. *R.temporaria*, a common garden frog is used intensively in evolutionary biology and ecology studies (Cano *et al.*, 2011) and mutations are very important in evolution (Akashi *et al.*, 2012). The current study shows that Parapinopsin is expressed in *X.laevis* and

X.tropicalis and is mutated in *R.tigrina* and *R.temporaria* at cDNA level and heterozygous in *R.temporaria* gDNA (wild type + mutated one). *R.tigrina* and *R.temporaria* are modern true frogs as compared to *X.laevis* and *X.tropicalis* and belong to the families *Ranidae* and *Pipidae* respectively (Roelants *et al.*, 2007), as Parapinopsin is not expressed in higher vertebrates.

Parapinopsin protein is a vertebrate primitive non-visual opsin protein involved in UV reception. In visual opsin proteins, e.g. rhodopsin a very important counterion, Glu, is present at position 113 and is involved in visual reception. The same counterion, Glu, is present in Parapinopsin at position 113, as well as at position 181, whereas invertebrate visual opsin has the counterion Glu at position 181. Hence Parapinopsin is a very important protein according to an evolutionary point of view between modern vertebrate and invertebrate opsins (Kawano-Yamashita *et al.*, 2011).

The current study revealed a point mutation in *R.temporaria* and *R.tigrina* (*Hoplobatrachus tigerinus*) Parapinopsin in terms of evolutionary genetics. Both are common true frogs and belong to family *Ranidae* which represent the evolution point where Parapinopsin may lose its function due to a point mutation which results in a stop codon and hence protein truncation (cDNA study results).

Chapter 4 Detection of Single Nucleotide Polymorphism by the use of Sequence Specific Primers using PCR (SSP-PCR)

Sequence specific primers in a PCR reaction are designed to single allele specific sequences, and result in amplification of a target sequence which completely matches the primer sequence. A pair of forward SSP-PCR primers has one or two 3' mismatches to the target sequence and amplifies the allele specific sequence in genomic DNA; hence SSP-PCR is used to discriminate between different alleles in a single PCR reaction. Taq DNA polymerase in a PCR reaction only recognizes and amplifies the completely matched 3' template end, as it lacks 3' to 5' exonuclease proofreading activity. SSP-PCR results are either positive or negative and can be easily detected by simple agarose gel electrophoresis.

4.1 Aim of Study

The aim of this study was to confirm a single nucleotide polymorphism (SNP) by SSP-PCR or a point mutation at the 2nd position of the 156th codon in *R.temporaria* and *R.tigrina* Parapinopsin, using gDNA. Two allele specific forward primers and one common reverse primer were designed to bind with the corresponding allele, and polymorphism was detected by PCR and electrophoresis.

Tadpoles of *R.temporaria* and brain of *R.tigrina* were used for the genomic DNA study. *X.laevis* gDNA and artificial heterozygotes were used as positive controls.

4.2 SSP- PCR Primer Designing

X.tropicalis PNP cDNA sequence was retrieved from NCBI. Four forward primers and one common reverse primer were designed against *X.tropicalis* PNP cDNA. All forward primers had a common sequence except 3' end mismatches that corresponded to the 2/156th codon. The sequence of *X.tropicalis* PNP cDNA is shown in table 4.1a and SSP-PCR primer sequences are given below in table 4.1 b. Forward primer set SSP1pnpA and SSP1pnpG have one 3' end mismatch while forward primer set SSP2pnpTA and

SSP2pnpTG have two 3'end mismatches *i.e.* at third last and last nucleotide; this will increase the primer specificity to bind to specific allele sequences. In the case of *R.temporaria*, which from sequence analysis may contain two different alleles “A” and “G” at the 2/156th codon, SSP-PCR primers should bind with their corresponding allele specific sequences only. Complete details of SSP-PCR primers are given in chapter 2 table 2.4.

a) <i>X.tropicalis</i> Parapinopsin cDNA Sequence with highlighted SSP-PCR primer positions (5'→3')		
<p>tgttgccctgacaatttcggtgactaaaaagtctcagaaccaaatggccgacgaggctcttctcccaccaatgat gaaatgtgaccaatgaagagatgcaccctgggaaggttctgatgccccggattggttacaccattctggccttaat aatggccgtattctgtgcccgcgtctcttctcacaagtcacggtgattgtggtgactttcaaatatcgccaact gcgccacccaatcaactactcgtcgtggtcaacctggccatcgccgaccttggcgtcacctgctgggcggggcctt gacagtggagacaaatgccgtgggctactttaacctggggagagtgggctgcgtcattgaggggttcgcgctggc atTTTTTggcatcgcggtttgtgcaccattgcagtaatagcgttggaccgagtggtcgtggtgtgcaagccaat gggcacactcacattcaccctccaaacaagcactggcgccgattgctgcttctctggatctggtcgctcatatggaa tacgcccccgctgttcggctgggtagctacgagctggaaggggtgatgacatcctgcgcccccaactgggtacag cgccgacccagtcacaatgtcctacatcgtctgctacttctccttctgctttgccatcccattcctcatcatcgt ggggtcctacgggtacctgatgtggactctgcggcAGGTTGCCAAGTTGGGTGTAGCGGAAGGAGGTACAACGAG TAAAGCCGAGGTGCAAGTCTCCCGTATGGTGATTGTCATGATCTTGGCTTTCTGGTCTGTTGGCTCCCGTACGC CGCCTTTGCCATGACGGTTGTGGCAAATCCCGAATGCACATTGACCCATTATAGCCACCGTACCCATGTACCT GACCAAAACCAGCACCGTCTATAATCCAATTATCTACATTTTCATGAACAAGCAGTTCCAAGAATGTGTCAATTCC CTTCTGTTCTGCGGGAGGAATCCTTGGGCTGCTGAAAAGTCAAGTTCTATGGAACTTCTATCAGCGTAACGAG TGGAACGCCTACGAAACGTGGCCAAGTGGCTCCGGCGTAACGTGCTTTTATCTCCAGTATAACCGG</p>		
b) SSP-PCR Primers & Sequence of <i>R.temporaria</i>		
Forward/Reverse Primer (5'→ 3')		<i>R.temporaria</i> gDNA sequence (181bp) (5'→ 3')
SSP1pnpA	cggcattgctgcttctctggatct a	<p>cggcattgctgcttctctggatcta/ggtcgctcat atggaatacgcccccgctgttcggctggggtagct acgagctggaaggggtgatgacatcctgcgcccc aactggtacagcgccgacccagtcacaatgtccta catcgtctgctacttctccttctgctttgccatcc cattcctc</p>
SSP1pnpG	cggcattgctgcttctctggatct g	
SSP2pnpTA	cggcattgctgcttctctggat tt a	
SSP2pnpTG	cggcattgctgcttctctggat ttg	
pnpR	gaggaatgggatggcaaagca	

Table4.1 *X.tropicalis* Parapinopsin cDNA sequence and SSP-PCR primer sequence

Part “a” of the above table shows the PNP cDNA sequence of *X.tropicalis*, where the red nucleotide “G” is polymorphic in *R.temporaria*. Start and stop codons are highlighted in pink and blue respectively. Lowercase letters show the cDNA sequence up to the start of exon3 and uppercase letters show the cDNA sequence of exons 3 & 4. Grey highlighted sequences show the position of SSP-PCR primers. In part “b” of the table the SSP-PCR primers are given with the amplified region of 181bp. All four forward primers differ at their 3' end nucleotides (shown in red). Forward primers 1 and 2 have one 3' mismatch while forward primers 3 and 4 have two mismatches at the 3' end.

4.3 SSP-PCR amplification of *X.laevis* Parapinopsin gDNA

gDNA was extracted from *X.laevis* whole tadpoles and quantified. A PCR reaction was carried out on the gDNA (200-300ng) of *X.laevis* using SSP1pnpA/ SSP1pnpG forward and pnp reverse primers. A 181bp PCR product would be expected to be produced with SSP1pnpG forward only. The PCR reaction results are shown in Figure 4.1.

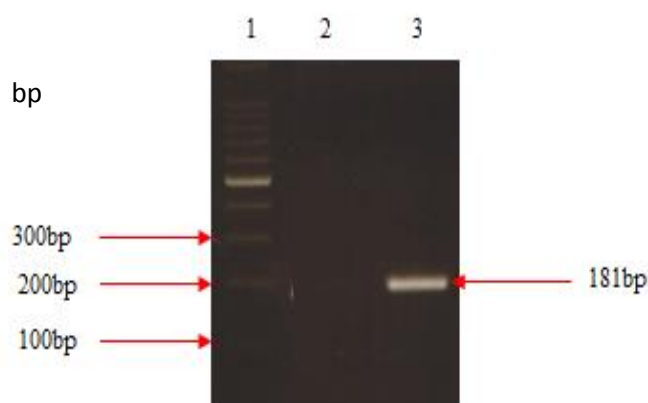


Figure 4.1 Electrophoretic analysis of PCR reaction using SSP1pnpA, SSP1pnpG forward and pnp reverse primers on *X.laevis* gDNA.

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: No PCR amplification with SSP1pnpA forward and pnp reverse

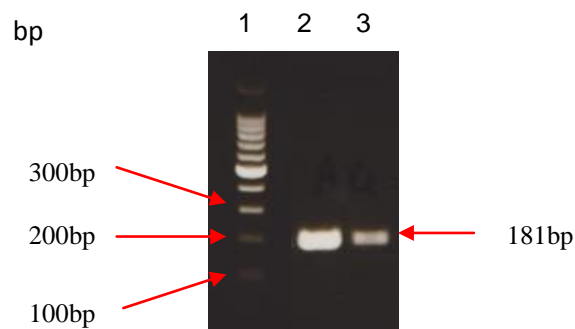
Lane 3: SSP1pnpG forward and pnp reverse PCR product

A 2% agarose gel was used to analyse the SSP1pnpA and SSP1pnpG PCR products through electrophoresis. PCR conducted on gDNA of *X.laevis* shows no amplification product appeared in lane 2 using SSP1pnpA while a 181bp band in lane 3 with SSP1pnpG indicates that both alleles of Parapinopsin gene in gDNA of *X.laevis* tadpoles contained “G” at 2/156th codon.

4.4 SSP-PCR Amplification of *Rana Temporaria* Parapinopsin gDNA

gDNA was extracted from *R. temporaria* and quantified. A PCR reaction was carried out on the gDNA (200-300ng) using SSP1pnpA/ SSP1pnpG and SSP1pnpTA/ SSP1pnpTG forward and pnp reverse primers. PCR reaction results are shown in Figure 4.2 conducted on gDNA of *R. temporaria*.

(a)



(b)

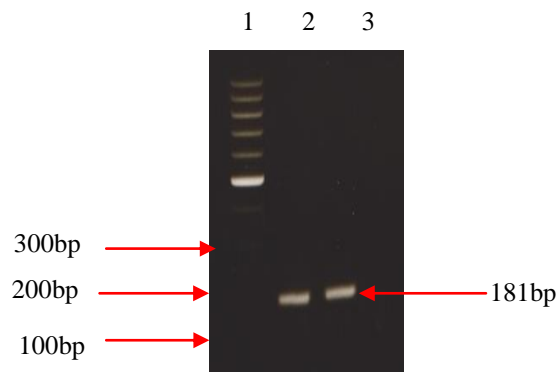


Figure 4.2 Electrophoretic Analysis of PCR products from *R. temporaria* gDNA with SSP1pnpA, SSP1pnpG, SSP2pnpTA, SSP2pnpTG forward primers and common pnp reverse primer

(a)

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: SSP1pnpA forward and pnp reverse PCR product

Lane 3: SSP1pnpG forward and pnp reverse PCR product

(b)

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: SSP2pnpTA forward and pnp reverse PCR product

Lane 3: SSP2pnpTG forward and pnp reverse PCR product

A 2% agarose gel was used to analyse the *R. temporaria* gDNA SSP1pnpA, SSP1pnpG, SSP2pnpTA and SSP2pnpTG PCR products through electrophoresis.

In lanes 2 and 3 of gel picture (a) and (b), the 181bp bands indicate that the Parapinopsin gene in gDNA of *R. temporaria* tadpoles is polymorphic at 2/156th codon.

4.5 SSP-PCR Amplification of *R. tigrina* Parapinopsin gDNA

gDNA was extracted from *R. tigrina* and quantified. A PCR reaction was carried out on the gDNA (200-300ng) using SSP1pnpA/ SSP1pnpG forward and pnp reverse primers. PCR reaction results are shown in figure 4.3.

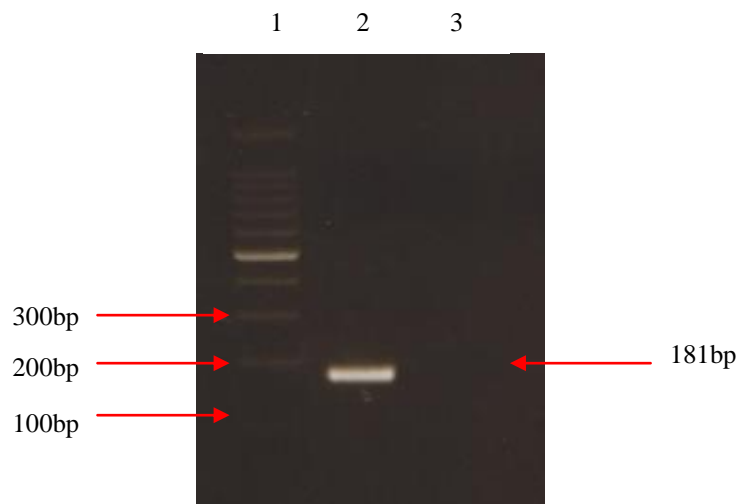


Figure 4.3 Electrophoretic analysis of PCR reaction using SSP1pnpA, SSP1pnpG forward and pnp reverse primers on *R. tigrina* gDNA.

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR amplifications with SSP1pnpA forward and pnp reverse

Lane 3: No PCR amplification with SSP1pnpG forward and pnp reverse PCR

A 2% agarose gel was used to analyse the SSP1pnpA and SSP1pnpG PCR products through electrophoresis.

The *R.tigrina* gDNA 181bp band in lane 2 with SSP1pnpA indicates that the both alleles of Parapinopsin gene in gDNA of *R.tigrina* contained “A” at 2nd position of 156th codon while no PCR amplification in lane 3 with SSP1pnpG.

4.6 SSP-PCR Amplification of Artificial Heterozygotes

Artificial heterozygotes were made by mixing cloned *X.laevis* and *R.temporaria* Parapinopsin cDNA in exactly the same ratio. *X.laevis* cDNA contains “G” while *R.temporaria* cDNA contained “A” at 2nd position of 156th codon. A PCR reaction was carried out using SSP1pnpA/ SSP1pnpG forward and pnp reverse primer. The 181bp PCR product size should be produced with SSP1pnpA as well as with SSP1pnpG forward due to the presence of heterozygosity at 2/156th codon. The PCR reaction results are shown in figure 4.4.

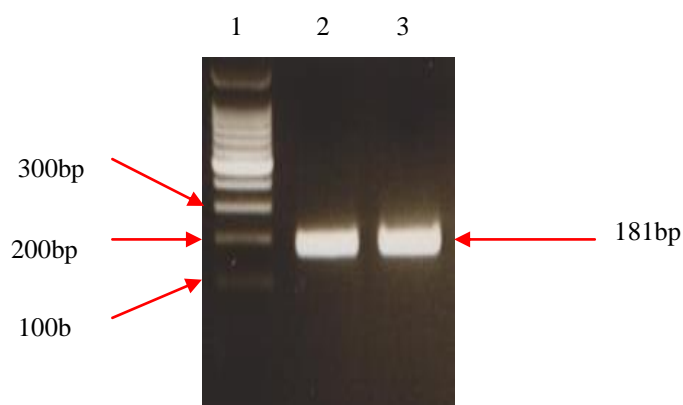


Figure 4.4 Electrophoretic analysis of the PCR reaction using SSP1pnpA, SSP1pnpG forward and pnp reverse primers on artificial heterozygotes of cDNA.

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR amplification with SSP1pnpA forward and pnp reverse

Lane 3: PCR amplification with SSP1pnpG forward and pnp reverse

Artificial heterozygotes were produced by mixing cloned cDNAs of *X.laevis* and *R.temporaria* Parapinopsin in exactly same ratios. A 2% agarose gel was used to analyse the SSP1pnpA and SSP1pnpG PCR products through electrophoresis.

The 181bp band in lanes 1 & 2 with SSP1pnpA and SSP1pnpG indicates that both alleles, “A” and “G” at the 2/156th codon, were present in the mixture.

4.7 Summary of SSP-PCR Results

Sample detail	PNP Genotype at 2 nd position of 156 th codon	Amplification with SSP-PCR forwards and common reverse primer	
		pnpG	pnpA
Artificial heterozygotes (Mixture of cloned cDNA of <i>R. temporaria</i> and <i>X. laevis</i>)	A/G	Yes	Yes
<i>X. laevis</i> gDNA	G	Yes	No
<i>R. temporaria</i> gDNA	A/G	Yes	Yes
<i>R. tigrina</i> gDNA	A	No	Yes

Table4.2 Summary of SSP-PCR primer results according to genotype of tadpoles

Column one of the table represents the gDNA of *X.laevis*, *R.temporaria* and *R.tigrina* tadpoles or cDNA mixtures of artificial heterozygotes used as a control during SSP-PCR study. The 2nd column shows the suggested genotype corresponding to 2nd position of 156th codon. The 3rd column shows the results of SSP-PCR primers. In the case of *X.laevis* gDNA, pnpG amplifies and pnpA amplifies *R.tigrina* gDNA. In the case of *R.temporaria* and the control artificial heterozygotes, both pnpA and pnpG primers bound with both alleles at 2/156th codon and gave amplification with both forward primers.

4.8 SSP-PCR Vs Sequencing Results

The Parapinopsin SSP-PCR results are in full accordance with the sequencing results giving amplification of allele specific bands by allele specific primers. A summary of both the results is given below in table 4.3.

Result summary of sequencing Vs SSP-PCR

<u>Sample information</u>	<u>Sequencing result</u>	<u>SSP-PCR results</u>		
	Genotype at 2nd position of 156th codon	Amplification with pnpG	Amplification with pnpA	Genotype
<i>X.laevis</i> gDNA	G	Yes	No	G
<i>R.temporaria</i> gDNA	A/G	Yes	Yes	A/G
<i>R.tigrina</i> gDNA	A	No	Yes	A
Artificial heterozygotes	A/G	Yes	Yes	A/G

Table4.3 Result summary of sequencing Vs SSP-PCR

Column one of the above table represents the gDNA of *X.laevis*, *R.temporaria* and *R.tigrina* tadpoles or cDNA mixtures of artificial heterozygotes used as a control during SSP-PCR study. The 2nd column shows the PNP genotype at 2/156th codon. The 3rd column shows the results using SSP-PCR primers.

Table 4.3 confirms the agreement between sequencing and SSP-PCR results and indicates a polymorphism at 2/156th codon in *R.temporaria* and a point mutation in *R.tigrina*.

4.9 Discussion

SSP-PCR (Strand Specific Primer- Polymerase Chain Reaction) is one of the important techniques used for SNPs genotyping. A single base pair mismatch at the 3' end of a primer is used for particular allele discrimination. This technique is preferred over RFLP (Restriction Fragment Length Polymorphism), RAPD (Random Amplified polymorphism Detection), Microarray and Sequencing *etc.* because it is cost effective and time saving. The PCR-SSP technique was first described by Olerup in 1991. Genotyping through SSP-PCR and agarose gel electrophoresis results in the presence or absence of an allele specific PCR product. Hence the SSP-PCR technique is used to determine particular alleles on a chromosome (Steinhardt *et al.*, 2009).

The current part of the project is based on SSP-PCR analysis of Parapinopsin gDNA from *X. laevis*, *R.temporaria* and *R.tigrina*. The confirmation of heterozygosity and particularly a point mutation is the key issue in the current work. Sequencing analysis of the Parapinopsin gene in the previous chapter revealed a point mutation from G → A in *R. temporaria/R. tigrina* Parapinopsin cDNA and SNP (Single Nucleotide Polymorphism) in *R. temporaria* gDNA at 2/156th codon. This point mutation is important as it creates a stop codon from UAG resulting in protein truncation. Sequencing indicated that two different alleles in *R. temporaria* Parapinopsin gDNA were present, one allele contained A and the other allele contained G. To confirm heterozygosity, two SSP-PCR forward primers pnpA and pnpG and a common reverse primer were designed. Forward primers pnpA and pnpG differ at their 3' ends having A and G respectively. Both forward primers will bind to their respective alleles. The SSP-PCR conditions were optimized for effective amplification with two forward primers and one common reverse primer.

Generally SSP-PCR primers have one mismatch at the 3' end, though sometimes a single base pair mismatch fails to discriminate between two alleles especially if it's due to low mutant to wild type sequence ratio. More than 1 mismatch at the 3' end and mismatches in the second to last nucleotide help to destabilize the 3' end and increase the chances of discrimination between two alleles (Kwok *et al.*, 1994). Hence, set 1

forward primer was SSP1pnpG and SSP1pnpA with one mismatch, and set 2 forward primers with 2 mismatches SSP2pnpG and SSP2pnpA were used during the current project.

The positive control *X. laevis* gDNA results showed an amplified band with SSP-PCR forward primer having G at its 3' end and common reverse primer and no amplification with SSP-PCR forward primers that have A at the 3' end indicating that both alleles of Parapinopsin 2/156th codon have G. The *R. tigrina* gDNA SSP-PCR results showed amplification with forward primers that end in A and no amplification with forward primers that end at G, and confirms a point mutation with both copies of the gene having A. The *R. temporaria* gDNA SSP-PCR result showed amplification with both sets of forward primers that have A as well as G at the 3' ends, confirming heterozygosity. One copy of the gene has A and other copy of gene has G.

The SSP-PCR primer sequence, primer concentration and annealing temperature are critical steps for successful and balanced amplifications of both alleles (Wu *et al.*, 2008, Talaat *et al.*, 2013).

To check the authenticity of the SSP-PCR technique, artificial heterozygotes were created (Richter *et al.*, 2013, Savage *et al.*, 1996, Othema and Sommer, 1993) , by mixing equal ratios of cloned cDNAs from *X. laevis* a positive control and *R. temporaria* A at 2/156th codon. Both SSP-PCR primers having A and G at their 3' ends amplified the alleles and two bands were seen on 2% agarose gel electrophoresis about equal in intensity. Thus the artificial heterozygote results were according to expectations: that both forward primers bind with, and amplify their respective allele. The SSP-PCR technique is overall helpful to confirm a point mutation, supports the Parapinopsin sequencing study results and strengthens the knowledge about Parapinopsin gene in different frog species.

Chapter 5 Study of Parapinopsin Protein

A Parapinopsin protein study was carried out on different frog species by Western Blotting.

5.1 Aim of Study

The aim of the study was to investigate the expression of Parapinopsin protein in *X. laevis*, *R. temporaria* and *R. tigrina*. Parapinopsin is a homodimeric protein and belongs to a group of 7 transmembrane G- protein coupled receptors (GPCR). In the previous chapter evidence has accumulated indicating a point mutation from G→A at 2nd position of 156th codon in both *R. temporaria* and *R. tigrina*. This mutation creates a stop codon from UGG→UAG right after the 155th amino acid positioned in the 4th helix. Hence a truncated protein or perhaps no protein will be made. Antibodies were made to regions of Parapinopsin to investigate the protein made by means of Western Blotting.

5.2 Study of Parapinopsin Protein: Peptide Design

The *X. tropicalis* Parapinopsin protein sequence reported at NCBI (NP_998830.1) was used to design peptides suitable for raising polyclonal antibodies for Western Blotting. Open reading frames (ORF) for *X. tropicalis*, *X. laevis* and *R. temporaria* / *R. tigrina* Parapinopsin cDNA were generated by using ORF graphical analysis tool available at NCBI website (www.ncbi.nlm.nih.gov/gorf/gorf.html) and it was found that the G→A point mutation creates a stop codon, and hence it could result in protein truncation leading to abnormal Parapinopsin production or no production in the species. ORF for *X. tropicalis* and *R. temporaria* are shown below in Table 5.1.

(i) *X. tropicalis* PNP cDNA sequence (NCBI, NM_213665.1)

tgttgccctgacaatttcggtgactaaaaagtctcagaaccaa**atgg**ccgacgaggctcttctccaccaatgatgaatgtgaccaatga
agagatgcacccctgggaaggttctgatgccccgattggttacaccattctggctttaataatggccgtattctgtgccgcccgtctctttct
caacgtcacgggtgattgtggtgactttcaaatatgccaaactgcgccaccaatcaactactcgtggtcaacctggccatcgccgacctt
ggcgtcacccgtgctgggcccgggcttgacagtgaggagacaaatgcgctgggctactttaacctggggagagtgggctgctgattgagg
tttcgctggcattttttggcatcgcggtttgtgaccattgcagtaatagcgttgaccgagtgctggtgtgcaagccaatgggca
cactcacattacccccaaacaagcactggccggcattgctgcttctggtatct**g**gtcgtcatatggaatagccccgctgttcggctg
gggtagctacgagctggaaggggtgatgacatcctgcgccccaaactggtacagcgccgaccagtcacatgtcctacatcgtctgta
cttctcttctgctttgccatccattctcatcatcgtgggttctacgggtacctgatgtggactctgcccaggttgccaagtgggtgtg
agcgggaaggaggtacaacagagtaaaagccgaggtgcaagtctccggtatgggtgattgtcatgatcttggcttctcgtgctgttggctccg
tacgcccgtttgcatgacggtttggtgcaaatcccggaatgcacattgacccattatagccaccgtacccatgtacctgacaaaacc
agcaccgtctataatccaattatctacattttcatgaacaagcagttccaagaatgtgtcattcccttctgtctgcccggaggaaatccttgg
gctgctgaaaagtcaagttctatggaacttctatcagcg**taa**cagtggaacgcctacgaaacgtggccaagtggctccggcgtaacg
tgcttttatctccagtataaccggc

(ii) *X. tropicalis*/ *X. laevis* ORF



Length: 348 aa

Accept Alternative Initiation Codons

44 **atgg**ccgacgaggctcttctccacca**atgatg**aatgtgaccaat
M A D E A L L P P M M N V T N
89 gaagag**atg**caccctgggaaggttctgat**atg**ccccgattggttac
E E M H P G K V L M P R I G Y
134 accattctggctttaata**atgg**ccgtattctgtgcccgcgctctc
T I L A L I M A V F C A A A L
179 tttctcaacgtcacgggtgattgtggtgactttcaaatatcgccaa
F L N V T V I V V T F K Y R Q
224 ctgcccacccaatcaactactcgtggtcaacctggccatcgcc
L R H P I N Y S L V N L A I A
269 gaccttggcgtcacctgtgctgggcccgttgacagtgagagaca
D L G V T V L G G A L T V E T
314 aatgcgctgggctactttaacctggggagagtggtgctgcatt
N A V G Y F N L G R V G C V I
359 gagggttttcgcccgtggcatttttggcctgcggctttgtgcacc
E G F A V A F F G I A A L C T
404 attgcagtaatagcgttgaccgagtgctgtggtgtgcaagcca
I A V I A L D R V F V V C K P
449 **atggg**ccacactcacattcacccccaaacaagcactggccggcatt
M G T L T F T P K Q A L A G I
494 gctgcttccctggatcgtgctgggcccgttgacagtgagagaca
A A S W I W S L I W N T P P L
539 ttcggctggggtgactacgagctggaaggggtg**atg**acatcctgc
F G W G S Y E L E G V M T S C
584 gcccccaactgggtacagcgccgacccagtcac**atg**tctacatc
A P N W Y S A D P V N M S Y I
629 gtctgctacttctccttctgctttgccatcccattcctcatc
V C Y F S F C F A I P F L I I
674 tgggggtcctacgggtacctg**atg**tggactctgcggcagggtgcc
V G S Y G Y L M W I L R Q V A
719 aagttgggtgtagcgggaaggaggtacaacgagtaaaagccgagtg
K L G V A E G G T T S K A E V
764 caagtctcccgat**atgg**gtgattgtc**atg**atccttggcttccgtgc
Q V S R M V I V M I L A F L V
809 tgttggctcccgtagcgcgcttggcc**atg**acggttggtgcaaat
C W L P Y A A F A M T V V A N
854 cccgga**atg**cacattgacccattatagccacgctaccc**atg**tac
P G M H I D P I I A T V P M Y
899 ctgacccaaaaccgacccgtctataatccaattatctacattttc
L T K T S T V Y N P I I T Y I F
944 **atg**aacaagcagttccaagaatgtgtctattcccttctgtctgc
M N K Q F Q E C V I P F L F C
989 gggaggaatccttgggctgctgaaaagtcaagttct**atg**gaaact
G R N P W A A E K S S M E T
1034 tctatcagcgtaacgagtggaacgcctacgaaacgtggccaagt
S I S V T S G T P T K R G Q V
1079 gctccggcg**taa** 1090
A P A *

R. temporaria ORF

a

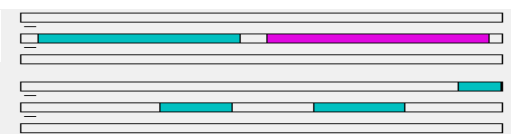


Length: 155 aa

Accept Alternative Initiation Codons

44 **atgg**ccgacgaggctcttctccacca**atgatg**aatgtgaccaat
M A D E A L L P P M M N V T N
89 gaagag**atg**caccctgggaaggttctgat**atg**ccccgattggttac
E E M H P G K V L M P R I G Y
134 accattctggctttaata**atgg**ccgtattctgtgcccgcgctctc
T I L A L I M A V F C A A A L
179 tttctcaacgtcacgggtgattgtggtgactttcaaatatcgccaa
F L N V T V I V V T F K Y R Q
224 ctgcccacccaatcaactactcgtggtcaacctggccatcgcc
L R H P I N Y S L V N L A I A
269 gaccttggcgtcacctgtgctgggcccgttgacagtgagagaca
D L G V T V L G G A L T V E T
314 aatgcgctgggctactttaacctggggagagtggtgctgcatt
N A V G Y F N L G R V G C V I
359 gagggttttcgcccgtggcatttttggcctgcggctttgtgcacc
E G F A V A F F G I A A L C T
404 attgcagtaatagcgttgaccgagtgctgtggtgtgcaagcca
I A V I A L D R V F V V C K P
449 **atggg**ccacactcacattcacccccaaacaagcactggccggcatt
M G T L T F T P K Q A L A G I
494 gctgcttccctggatc**taa** 511
A A S W I *

b



Length: 172 aa

Accept Alternative Initiation Codons

572 **atg**acatcctgcgcccccaactggtacagcgccgacccagtcac
M T S C A P N W Y S A D P V N
617 **atg**tctacatcgtctgctacttctctctctgcttggcattccca
M S Y I V C Y F S F C F A I P
662 ttcctcatcactggtgggtctcactggtacacgt**atg**tggactctg
F L I I V G S Y G Y L M W T L
707 cggcagggttgccaagtgtgggtgtagcgggaaggaggtacaacgag
R Q V A K L G V A E G G T T S
752 aaagccgaggtgcaagtcctccgt**atg**gtgattgtc**atg**tcttg
K A E V Q V S R M V I V M I L
797 gcttctcgtgtcgtgctccgtagcgcgcccgttggcattgacg
A F L V C W L P Y A A F A M
842 gttgtggcaaatcccgga**atg**cacattgacccattatagccacc
V V A N P G M H I D P I I A T
887 gtaccc**atg**tacctgacccaaaccagcagcgtctataatccaatt
V P M Y L T K T S T V Y N P I
932 atctacattttc**atg**aacaagcagttccaagaatgtgtcattccc
I Y I F M N K Q F Q E C V I P
977 tttctgtctcgtgggaggaatccttggctgctgcaaaagtcaagt
F L F C G R N P W A A E K S S
1022 tct**atg**gaaactctatcagcgtaacgagtggaacgcctacgaaa
S M E T S I S V T S G T P T K
1067 cgtggccaaagtgggtcccggt**taa** 1090
R G Q V A P A *

Table 5.1 *X. tropicalis* Parapinopsin cDNA sequence and Open Reading Frame (ORF) information (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>)

Part (i) of the above table shows the cDNA sequence of *X. tropicalis* Parapinopsin with highlighted start and stop codons in green and pink colour respectively, while bold red colour nucleotide “g” is a nucleotide of interest, mutated in *R. temporaria* from G→A creating a stop codon. In part (ii) of above table, column 1 represents the ORF of *X. tropicalis* / *X. laevis* with the longest stretch of nucleotide sequence with start and stop codon which on translation yields Parapinopsin of 348 amino acids. Column 2 represents the two ORF “a” and “b” for *R. temporaria*, due to a point mutation a stop codon is created at the 156th codon hence a truncated protein of 155 amino acids may be made.

5.2.1 Parapinopsin Sequence Study and Selection of Epitopes for the Generation of Polyclonal Antibodies

The sequence of *X. tropicalis* Parapinopsin 37.9kDa protein which comprises 348 amino acids (reported at NCBI) was used to design the peptide regions for antibody production. It was decided to make two antibodies one at the N-terminus and one at the C-terminus. The reason for selection of two antibodies was a point mutation found during the current project of sequencing study. A point mutation in Parapinopsin at the cDNA level of *R. temporaria* and *R. tigrina* creates a stop codon resulting in a truncated protein of 155 amino acids (molecular size of 16.55kDa). A stretch of 21 amino acids is missing between the stop and next potential start codon shown between red arrows in figure 5.1 below. Antibodies binding at the N and C termini would identify the protein by Western Blotting. It was assumed to get full intact protein of ~38kDa in the case of *X. laevis* with both N and C terminal antibodies. With *Rana* (*R. temporaria*/*R. tigrina*) if a truncated protein is made, then the antibody to the N terminus should bind because the N terminus is common to normal Parapinopsin and the truncated form. The C terminus antibody should bind to normal Parapinopsin but not the truncated form.

>*X. tropicalis* Parapinopsin Sequence (NCBI reference # 998830.1)

MADEALLPPMMNVTNEEMHPGKVLMPRIGYTILALIMAVFCAAALFNVTVIVVTFKYRQLRHPINYSLVNLAIAD
LGVTVLGGALTVEITNAVGYFNLGRVGCVIEGFAVAFFGIAALCTIAVIALDRVFVCKPMGTLTFTPKQALAGIAAS
WIWSLIWNTPLFGWGSYELEGVMTSCAPNWYSADPVNMSYIVCYFSFCFAIPFLIIVGSYGYLMWTLRQVAKLG
VAEGGTTSKAEVQVSRMVIVMILAFVLCWLPYAAFAMTVVANPGMHIDPIIATVPMYLTSTVYNPIIYIFMNKQ
FQECVIPFLFCGRNPWAAEKSSMETSSISVTSGTPTKRGQVAPA

Figure 5.1 *X. tropicalis* Parapinopsin protein sequence and selection of peptide regions

The figure shows *X. tropicalis* Parapinopsin protein of 348 amino acids, sequence reported at NCBI. The grey highlighted regions in between blue arrows of 12 amino acids at N and C termini, represent the peptide regions selected for Parapinopsin antibodies. The 21 amino acids region between red arrows from 156th-176th is missing in *R. temporaria* due to creation of a stop codon right after 155th amino acid. Hence the possible creation of two truncated proteins of 16.55kDa (1-155th amino acid) and 172 amino acids of 19.01kDa (from 177th to 348th amino acid).

5.2.2 Secondary Structure Prediction of Parapinopsin Protein

The Parapinopsin protein secondary structure was predicted by using software PSIPRED V3.3 for *X. tropicalis* / *X. laevis* as both have 100% sequence homology. The predicted secondary structure of Parapinopsin protein indicates that the protein comprises 7 helices shown in numbers 1 to 7 in bold blue colour below in figure 5.2. The point mutation which creates a stop codon was found in the 4th helix (labelled in red). Peptide regions of 12 amino acids which are highly specific to *X. tropicalis* Parapinopsin were selected at N and C termini for antibody production from 57th → 68th and 316th → 327th amino acids respectively, highlighted in green arrows. The peptide region for the N terminal antibody was found in a coil region between helix 1 and helix 2 as well as in helix 1 and the peptide region selected for C terminal antibody is positioned in the last coil region after the 7th helix, highlighted by a green arrow in

figure 5.2.

5.2.3 Production of Polyclonal Antibodies and Western Blot analysis

X. tropicalis Parapinopsin protein sequence was studied in detail and highly specific peptide regions with high antigenicity were selected. The selected peptide regions of 12 amino acids at N and C termini were unique to *X. tropicalis* Parapinopsin protein and this was confirmed by Protein- BLAST (Basic Local Alignment search Tool). Parapinopsin Polyclonal antibodies were produced by Generon, raised in rabbits, and used against frog protein extracts for Western Blotting analysis.

5.2.3.1 Antigenic Epitope Selection

Peptide regions were selected by following important rules, given in Table 2.6 by using “protein analysis tool/peptide antigen design tool for antibodies”

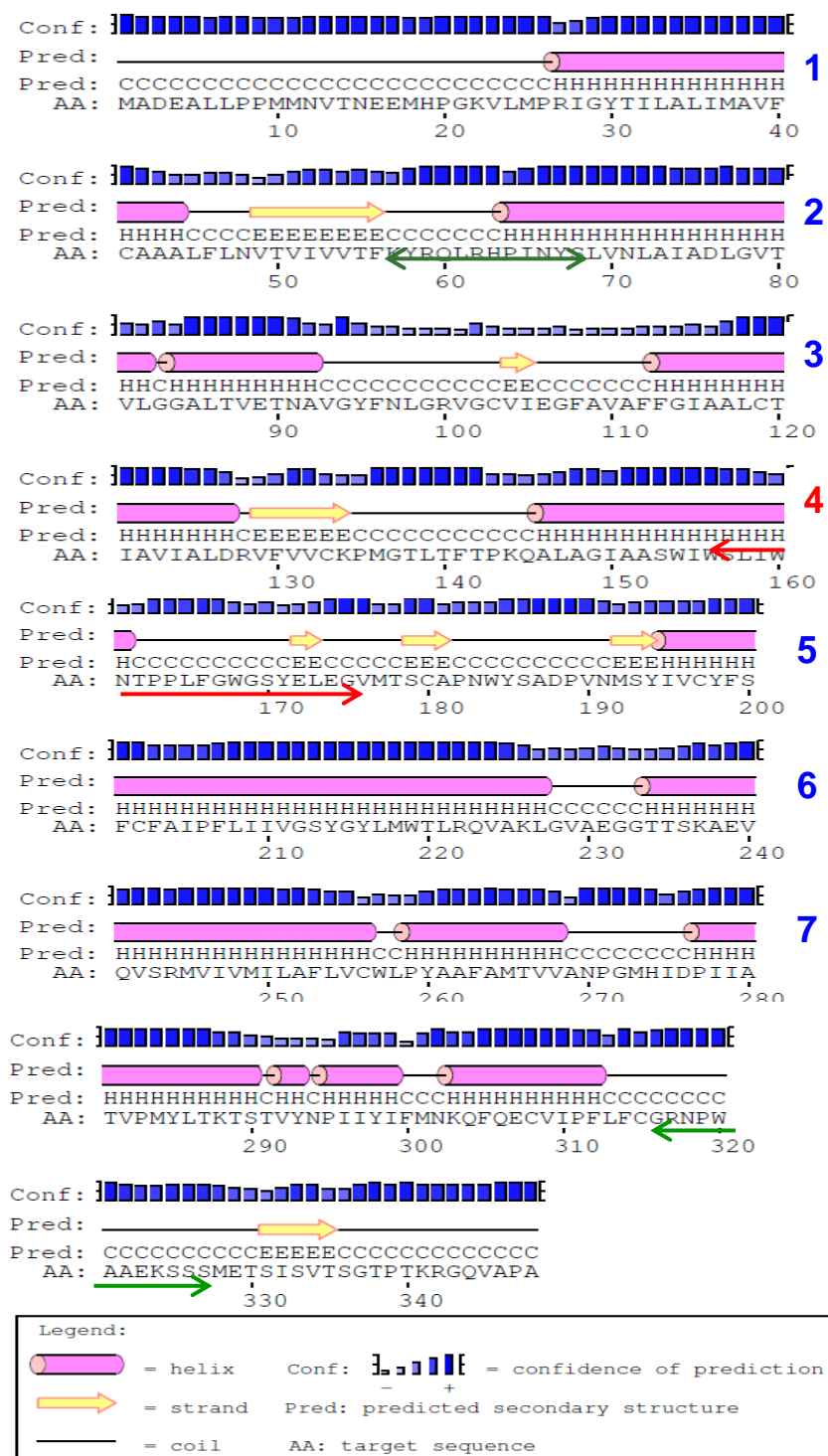


Figure 5.2 Predicted secondary structure of *X. tropicalis*/ *X. laevis* Parapinopsin protein (<http://bioinf.cs.ucl.ac.uk/psipred>)

The figure shows Parapinopsin protein secondary structure predicted by using software PSIPRED V3.3 of *X. tropicalis* / *X. laevis*. Parapinopsin protein comprises 7 helices shown in numbers 1 to 7 in bold blue colour. In the above figure **H=helix**, **E=strand**, **C=coil**. A point mutation which creates a stop codon and results in protein truncation was found in 4th helix underlined with a red arrow. Highly specific peptide regions of 12 amino acids with high antigenicity at N and C terminal were

selected for antibodies from 57th → 68th and 316th → 327th respectively, highlighted by green arrows.

5.2.4 Parapinopsin Protein 3D Structure Prediction

The *X. tropicalis* Parapinopsin 3D structure was predicted by using software **YASARA** (Yet Another Scientific Artificial Reality Application). It was found that Parapinopsin exists as a transmembrane helical protein like other opsins e.g. rhodopsin. The 3D structure of *X. tropicalis* Parapinopsin is shown in Figure 5.3.

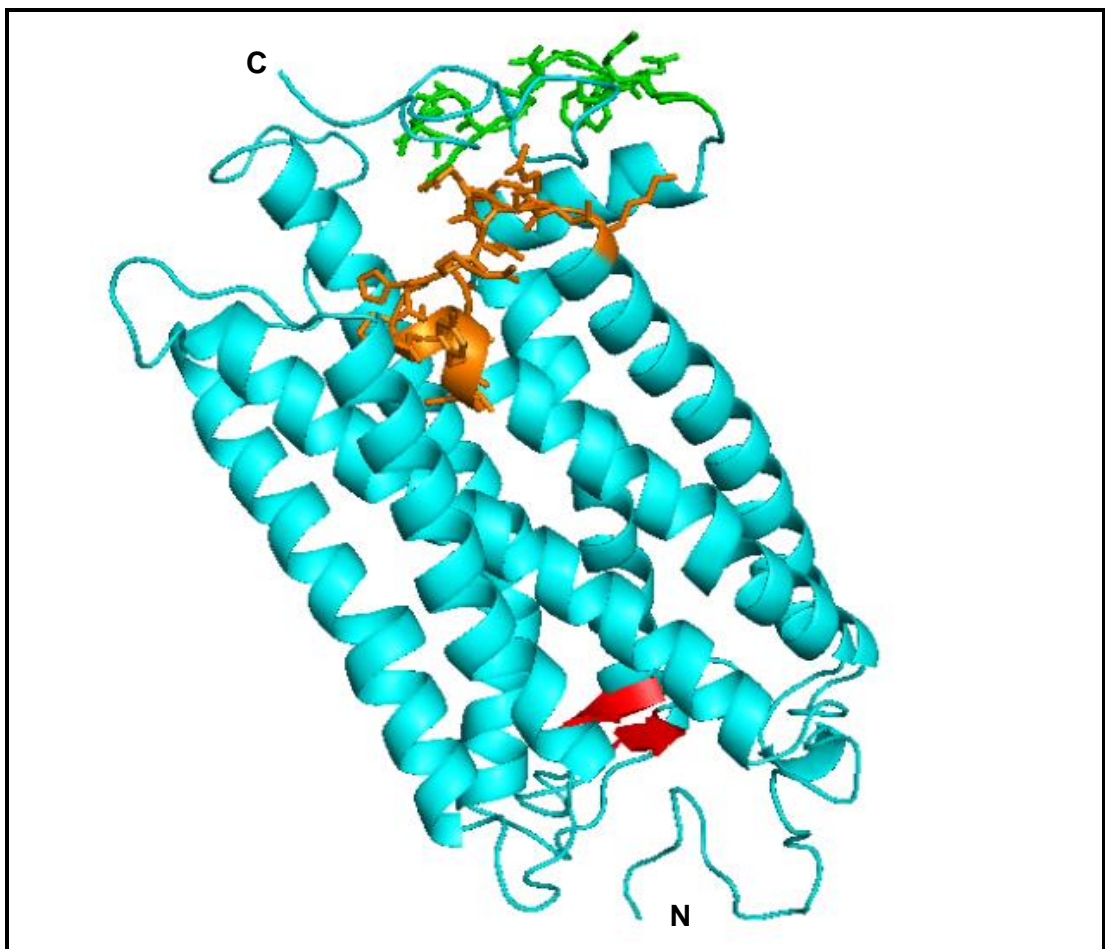


Figure 5.3 Predicted 3D structure of *X. tropicalis* Parapinopsin protein with selected N and C terminal peptides

(<http://www.yasara.com/index.html>)

The figure shows a ribbon presentation of the 3D structure of *X. tropicalis* Parapinopsin. Parapinopsin is a seven transmembrane helical protein. The “N” and “C” termini are labelled. The helices of the protein are shown I-VIII. The loop regions are shown in green, and β strands are shown in red. The peptide on the N terminus is shown in orange and the peptide on the C-terminus is in green.

5.2.5 *R. temporaria* Predicted Parapinopsin Protein 3D Structure

The normal, wild type, Parapinopsin protein exists as a 7 transmembrane protein where the 2nd extracellular loop region, between helix 4 and helix 5, is modified into β - strands, but in the case of *R. temporaria* the predicted structure of 3D Parapinopsin shows the absence of β - strands due to a point mutation which creates a stop codon (found at *R. temporaria* cDNA level). The predicted 3D structure of *R. temporaria* Parapinopsin protein by YASARA, with highlighted N and C termini peptides is shown in Figure 5.4.

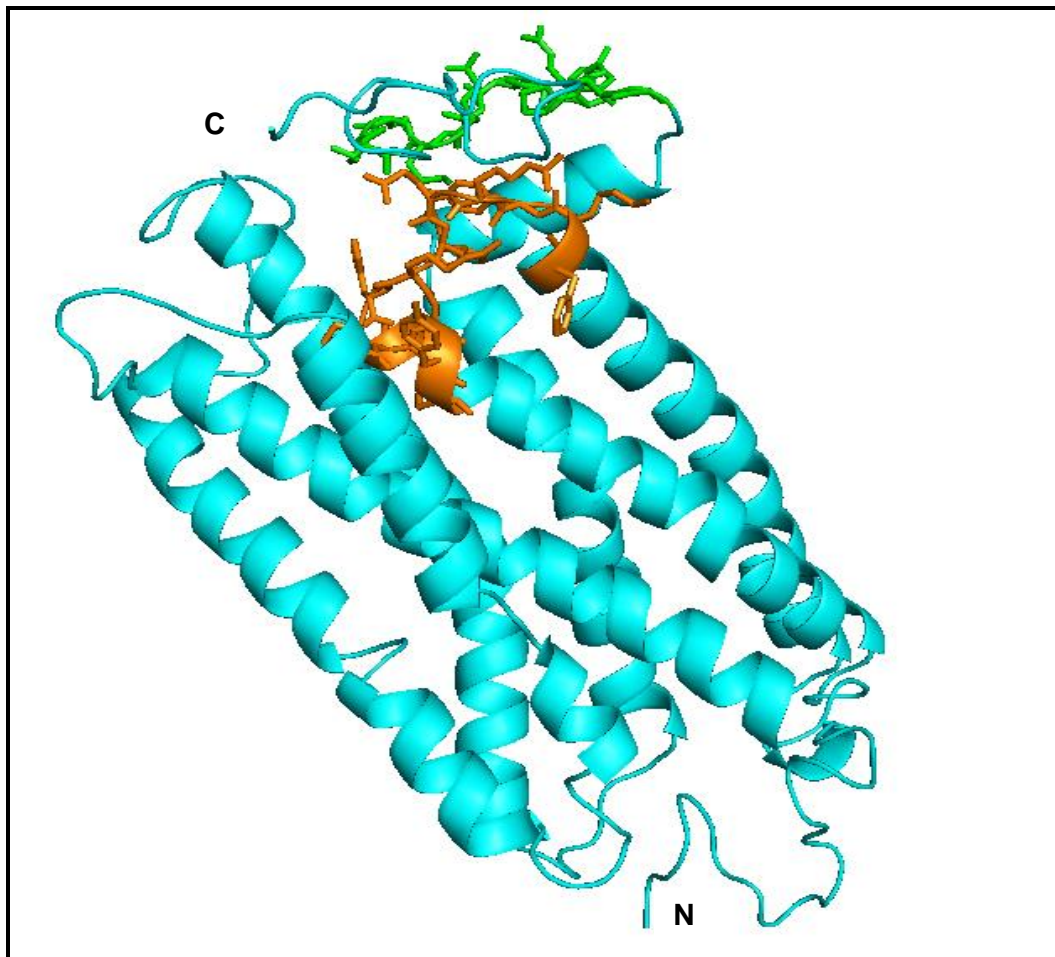


Figure 5.4 The predicted *R. temporaria* Parapinopsin protein

The above figure presents *R. temporaria* Parapinopsin 3D structure with highlighted N and C terminal peptides shown in orange and green respectively. The alpha helices and coil regions are shown in blue. Beta strands are missing in the *R. temporaria* Parapinopsin.

5.3 Gel staining and Parapinopsin protein analysis

X. laevis protein extract along with rat myofibrils and haemoglobin purified protein were used as controls for Western Blot analysis. Protein samples were loaded (20µg/lane, protein concentration details are given in Chapter 2 under section 2.17 and 2.18) in duplicate on 15% SDS gels; one gel for staining and one for blotting. The gels were stained, and conditions were optimized by using control protein samples along with frog protein samples prior to blotting. The 42kDa alpha actin in rat myofibril and 15kDa purified haemoglobin protein samples were used as control and loaded along with *X. laevis* and *R. temporaria* protein samples. Parapinopsin (~38kDa) or truncated Parapinopsin protein (17-19kDa) should be seen in the stained gel between two control protein samples which could be confirmed by using Parapinopsin specific antibodies later on. The stained gel Pictures 5.5 a and b are shown below.

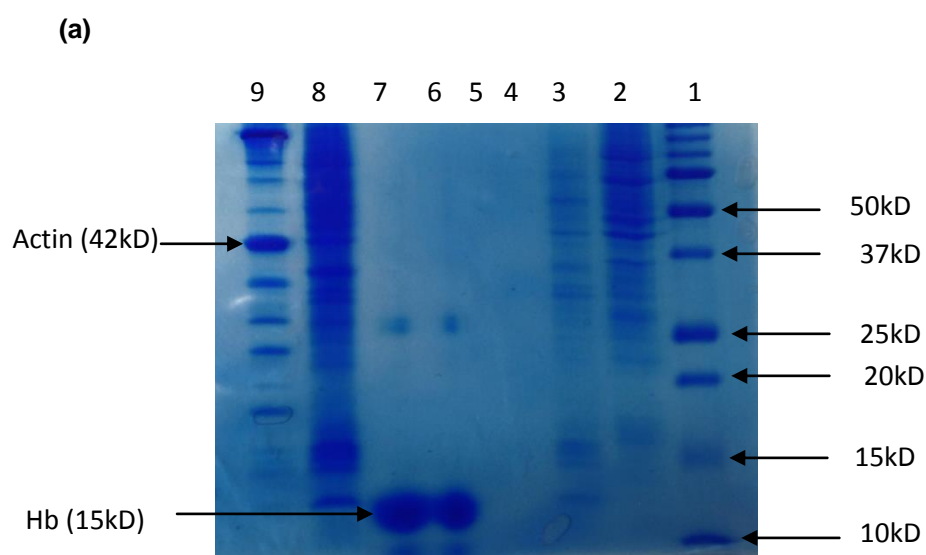


Figure 5.5 (a) 15% SDS gel electrophoretic analysis of *X. laevis*, *R. temporaria* Parapinopsin, actin and haemoglobin protein samples

A 15% SDS gel was used to analyse the protein samples through electrophoresis.

Lane 1: BioRad protein marker

Lane 2: *X. laevis* protein sample

Lane 3: *R. temporaria* protein sample 1

Lanes 4&5: empty

Lanes 6&7: Purified haemoglobin protein sample

Lane 8: *R. temporaria* protein sample 2

Lane 9: rat myofibril protein sample for actin

Lane 3 & 8 showed *R. temporaria* protein samples from different tadpoles. Parapinopsin protein should be found between the two positive control proteins and near to the 37kDa or 17kDa protein marker bands, in the case of a full or a truncated protein.

(b)

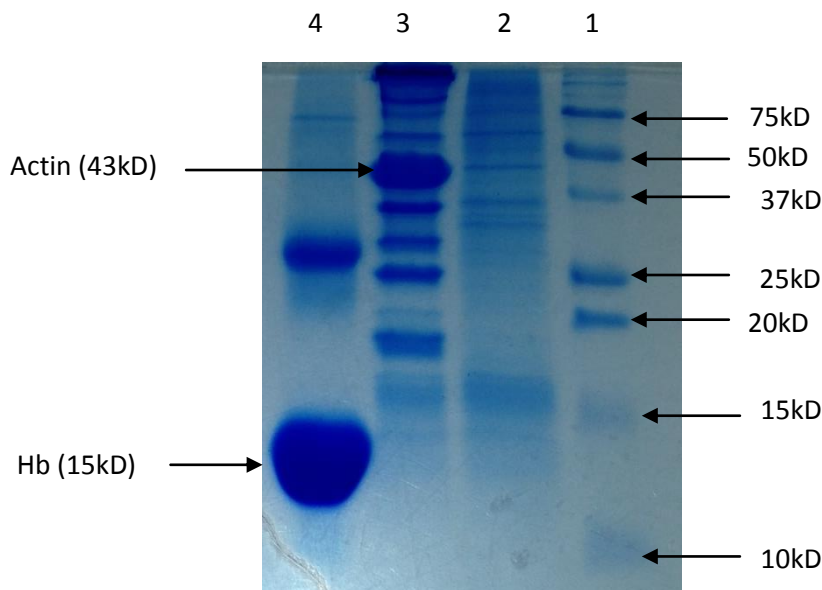


Figure 5.5 (b) 15% SDS gel electrophoretic analysis of *R. tigrina* Parapinopsin, actin and haemoglobin protein samples

A 15% SDS gel was used to analyse the protein samples through electrophoresis.

Lane 1: BioRad protein marker

Lane 2: *R. tigrina* protein sample

Lane 3: Rat Myofibrils protein sample

Lane 4: Haemoglobin purified protein

Lane 1 shows the protein marker. Lane 2 shows the *R. tigrina* brain protein sample. Lane 3 & 4 shows protein samples of rat myofibril for Actin protein and purified haemoglobin protein respectively, used as positive controls. Parapinopsin protein, if expressed, should be found between two positive control proteins and near to the 37.9kD or 19kDa protein marker bands, in the case of a full or a truncated protein.

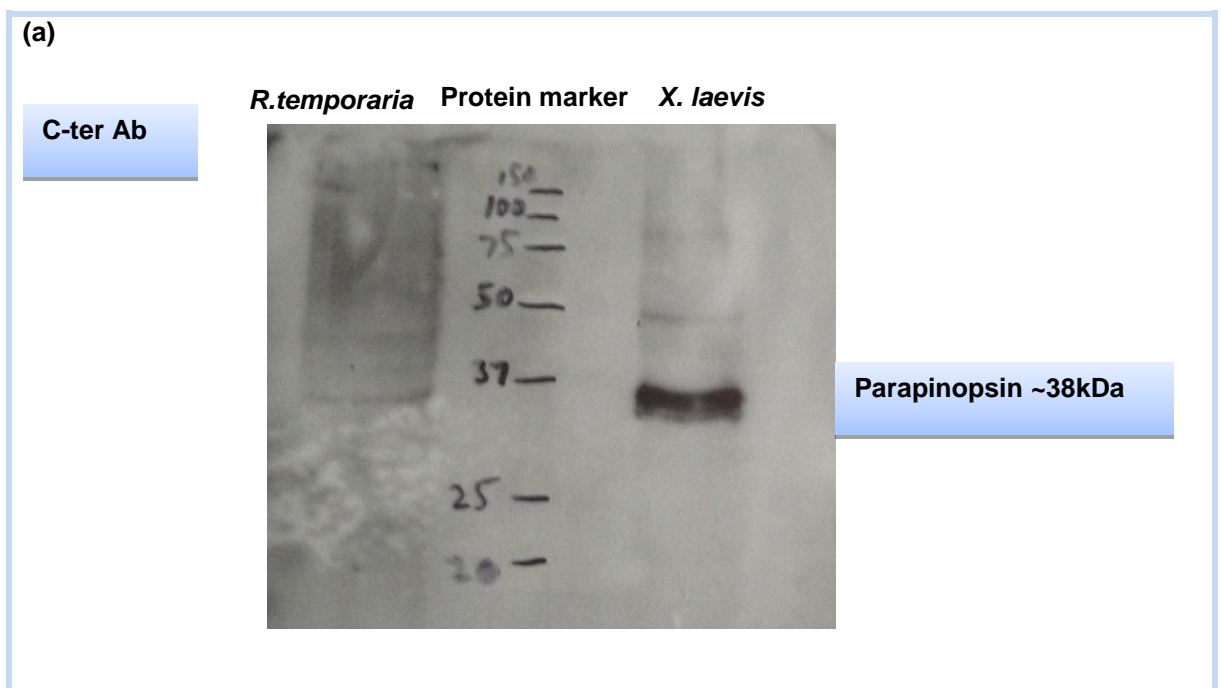
5.4 Western Blot analysis for *X. laevis* & *R. temporaria*

Parapinopsin Protein using N and C Terminal Antibodies

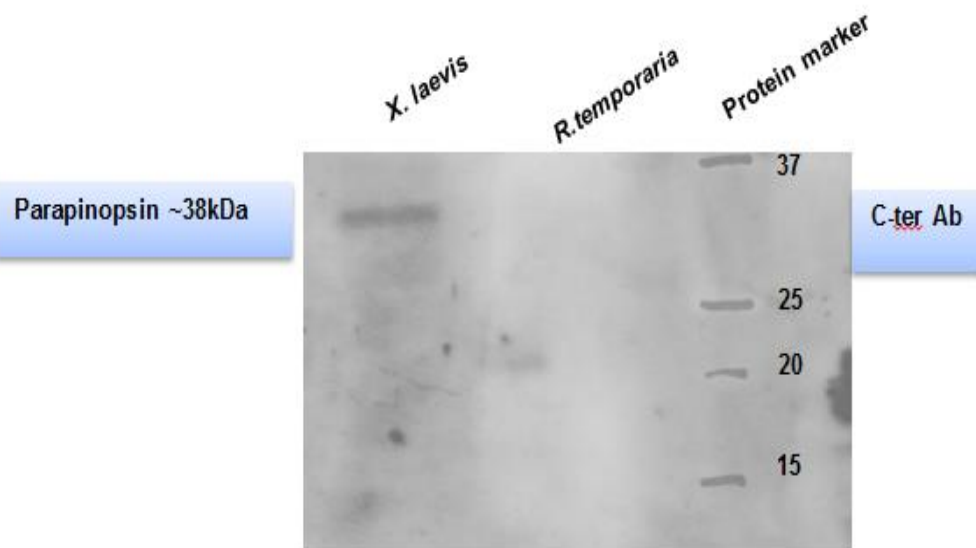
R. temporaria protein extract from whole tadpoles was used for Western Blot analysis.

Pineal brain tissues protein extract of adult *X. laevis* frogs was used as a positive control.

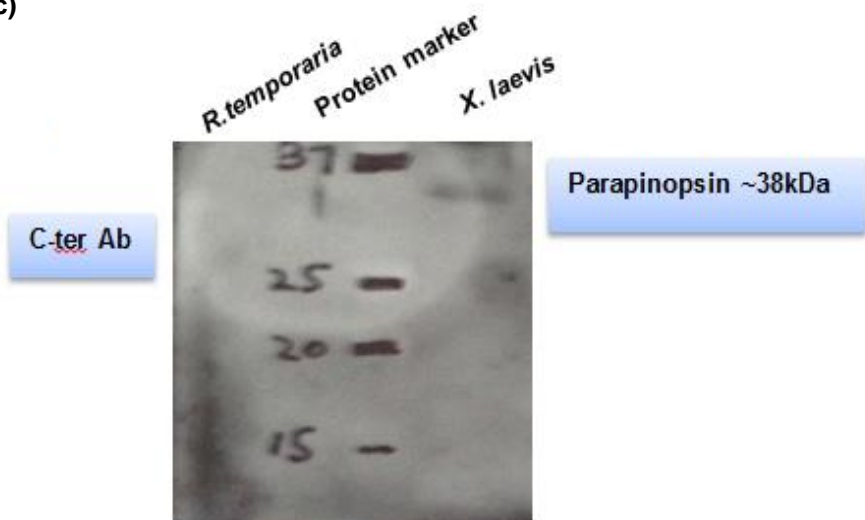
The same protein concentration and loading of both samples were used, and Western Blot results of *R. temporaria* and *X. laevis* with N and C terminals Parapinopsin antibodies are shown below in Figure 5.6



(b)



(c)



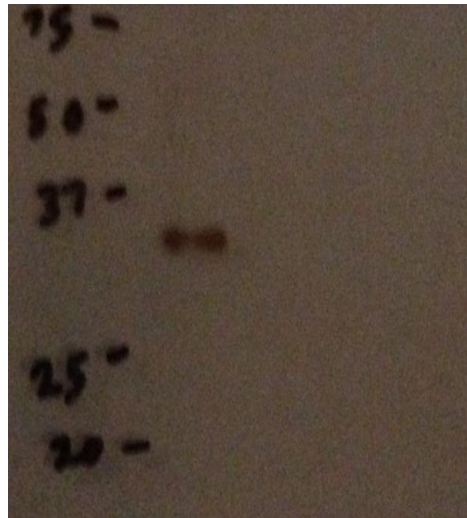
(d)

Protein
marker

X. laevis

R. temporaria

N-ter Ab



Parapinopsin ~38kDa

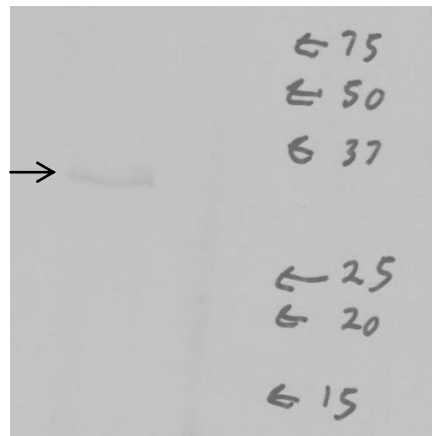
(e)

X. laevis

R. temporaria

Protein marker

Parapinopsin ~38kDa



N-ter Ab

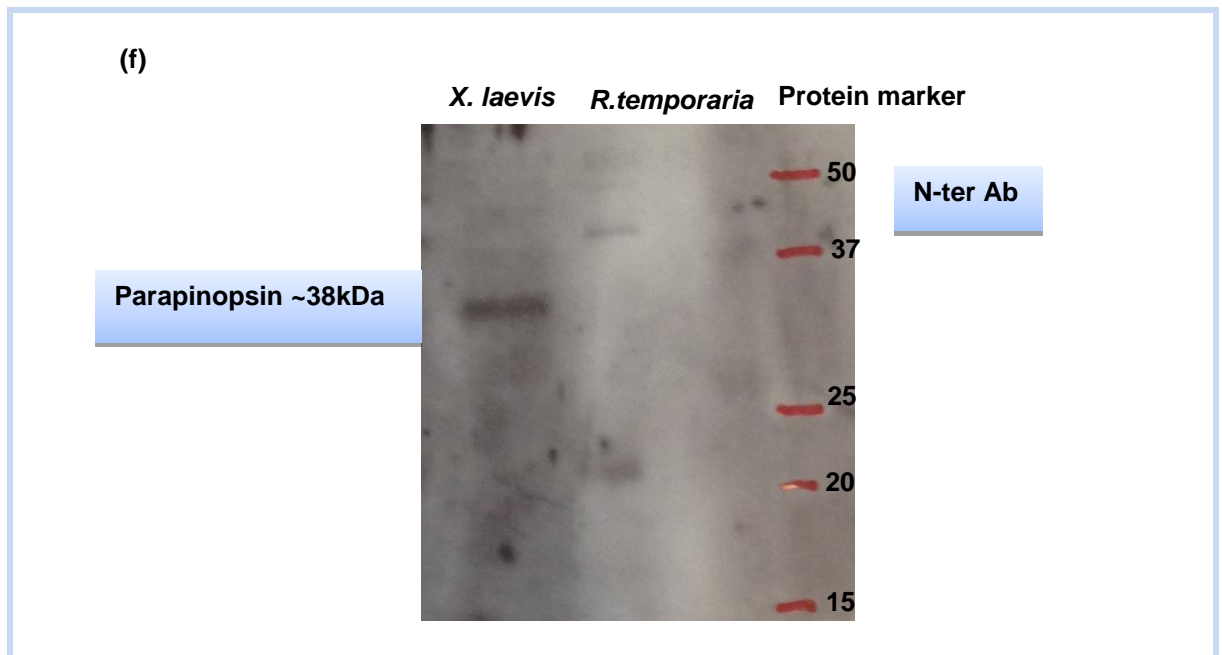


Figure 5.6 Western Blot analysis for *R. temporaria* & *X. laevis* Parapinopsin protein with C and N- terminal PNP antibodies

The above Figures illustrate the Western Blot analysis for Parapinopsin protein in *R. temporaria* tadpoles and *X. laevis* (positive control) pineal protein extracts.

C-terminal Antibody results with *X. laevis* & *R. temporaria*

(a) Figure shows 38kDa band with *X.laevis* and very faint band with *R. temporaria* at the same position

(b) A repeat of whole experiment shown in (a) with new protein sample

(c) A repeat of whole experiment shown in (b) with new protein sample N-terminal antibody results with *X. laevis* & *R. temporaria*

(d) Figure shows 38kDa band with *X.laevis*

(e) A repeat of whole experiment shown in (d) with new protein sample

(f) A repeat of whole experiment shown in (e) with new protein sample

The Western Blot analysis results shows that Parapinopsin protein is expressed in *X. laevis* and both N and C termini antibodies bind with ~38kDa Parapinopsin protein expressed in *X. laevis*. Both N and C terminal antibodies showed negative results with *R. temporaria* as there is no protein at 38kDa or at 17-19kDa positions.

5.4.1 Western Blotting analysis for *X. laevis* & *R. tigrina* Parapinopsin Protein using N and C terminal Antibodies

R. tigrina brain tissue protein extract was used for Western Blot analysis. Blotting results with N and C terminal antibodies are shown below in Figure 5.7.

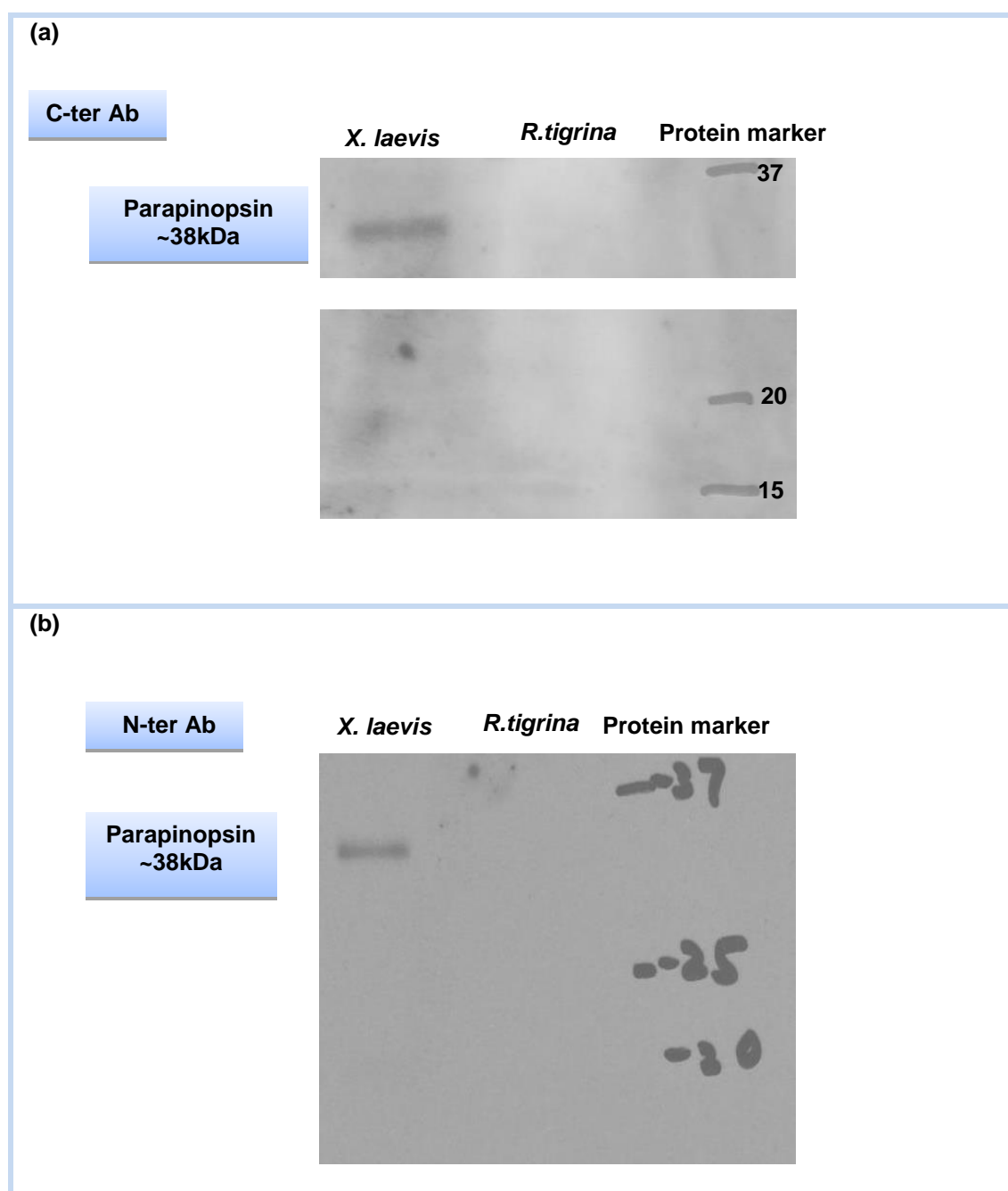


Fig 5.7 Western Blot analysis for *R. temporaria* & *X. laevis* Parapinopsin protein with C and N- terminal Parapinopsin antibodies

The figure illustrates the Western Blot analysis of Parapinopsin protein with *R. tigrina* adult frog brain and *X. laevis* (positive control) pineal protein extracts. In Part (a) of above figure the C-terminal Antibody shows the Western Blot result with C-terminal Parapinopsin antibody and in (b) the N terminal Antibody shows results with N-terminal Parapinopsin antibody.

The Western Blot analysis results show that Parapinopsin protein is only expressed in *X. laevis* and both N and C termini antibodies bind with Parapinopsin protein. Both N and C

terminal antibodies showed negative results with *R. tigrina* as there is no protein with ~ 17-19kDa band.

5.5 Discussion

Parapinopsin is expressed in cold-blooded vertebrates e.g. Catfish, Zebrafish, Lamprey, *X. tropicalis*, in the pineal organ (Wada *et al.*, 1998; Mano *et al.*, 1999) and is very important for their biological clock and UV reception (Shichida and Yamashita 2003, Yamashita 2007). There is also a recent report regarding expression of Parapinopsin in reptile, the Iguana parietal eye by Wada *et al.*, 2012.

The only reported sequence of Parapinopsin protein at NCBI in frogs is that of *X. tropicalis*. For Western Blot analysis Parapinopsin protein was studied in detail as no information about Parapinopsin protein structure is available. The Parapinopsin protein primary structure was predicted by ORF graphical analysis tool (www.ncbi.nlm.nih.gov/gorf/gorf.html). The secondary and 3D structure Models were predicted by using softwares PSIPRED V3.3 and YASARA respectively. YASARA (Yet Another Scientific Artificial Reality Application) is a modern versatile software most commonly used by Bioinformaticians (Alanazi *et al.*, 2011, Krieger E *et al.*, 2009, Venselaar *et al.*, 2010 and Dewi & Fatchiyah, 2013) and has many applications. YASARA was used during the current study on Parapinopsin protein structure prediction. In, *In Silico* protein modelling, a short stretch of conserved amino acid sequences called a “pattern” is used to predict the protein structure (Michał Brylinski *et al.*, 2006). YASARA predicts Parapinopsin 3D structure by using Rhodopsin protein as a template, a most studied protein (Lagerstrom and Schioth, 2008), where Parapinopsin showed 60% similarity with the target Rhodopsin (Tsukamoto *et al.*, 2009, Shichida and Matsuyama, 2009).

Sequence analysis during the current project revealed novel findings. 100% homology was found between Parapinopsin cDNA sequence of *X. tropicalis* and *X. laevis*. This is supported by Hirsch *et al.*, 2002, due to the species being close relatives, both shared most conserved sequences. ~99.9% cDNA sequence similarity was found between *X.*

tropicalis, *R. temporaria* and *R. tigrina*. The 0.1% dissimilarity is due to a point mutation from G→A at the 2nd position of the 156th codon which creates a stop codon UGG→UAG in *Rana*, instead of tryptophan, and result in protein truncation. A stop codon created in the middle of Parapinopsin, 348 amino acids (37.9 kDa) divides the protein into two halves amino acids 1- 155 (16.55 kDa) and amino acids 177-348 (19.01 kDa). A stretch of 21 amino acids is missing after the stop codon and the possible start codon of a second part of the protein. All this *In Silico* information about Parapinopsin protein was gathered by using Bioinformatics tools ORF finder and PSIPRED v3.3 available at NCBI and UCL websites and was used to design antibodies for Western Blot analysis.

The Opsin protein consists of a seven pass transmembrane helical structure joined with three cytoplasmic and three extracellular loops (Y. Shichida & T. Matsuyama, 2009; Lagerstrom and Schioth, 2008). The chromophore, 11 cis retinal, covalently binds through a Schiff base with L-lysine 296, present in helix seven (Koyanagi and Terakita, 2013). On light reception, the Schiff base becomes protonated and this protonation is stabilized by a negative charge on glutamic acid counterion at position 113. In Parapinopsin, in addition to counterion 113, another counterion glutamic acid at position 181 is also present in the 2nd extracellular loop between 4th and 5th helix. The 2nd extracellular loop is modified into β3 and β4 strands and plays an important role in binding of the chromophore. A water molecule helps in the formation of hydrogen binding around the Schiff base with the help of the carboxylic acid side chain of Glutamic acid (181) present in the β3 strand. A mutation at this position greatly affects the absorption properties of the chromophore (Shichida and Takahiro Yamashita 2003, Yamashita *et. al.*, 2011).

Phylogenetic studies show that primitive opsin has a counterion for Schiff base at position 181 like invertebrate opsins. The modern opsins have evolved a counterion at a new position for Schiff base at position 113 (Lamb, 2009), while Parapinopsin has glutamic acid at both positions. Therefore Parapinopsin is an important evolutionary link among vertebrate and invertebrate GPCR proteins (Terakita, 2005).

A point mutation which creates a stop codon in *R. temporaria* and *R. tigrina* protein is present in the 4th helix, hence the 2nd extracellular loop is missing in *R. temporaria* and *R. tigrina*. A potential new start codon is present after 21 amino acids in helix 5. β 3 and β 4 strands are missing in the predicted *R. temporaria* 3D Parapinopsin protein by YASARA when compared with the Bovine Rhodopsin protein model (www.yasara.org).

Polyclonal antibodies were produced by selecting common peptide regions from Parapinopsin in all three frogs' *X. laevis*, *R. temporaria* and *R. tigrina* to use as antigens. It was decided to make two antibodies one at the N-terminus and one at the C-terminus. The peptides were selected on the basis of high immunogenicity, and specificity was found with the help of BLASTp (basic local alignment search tool for protein) available at NCBI. Antibodies were raised in rabbits. Both antibodies should bind to intact Parapinopsin found in *X. laevis* and *X. tropicalis*. The antibody to the N-terminus should bind to the truncated form of Parapinopsin (17-19 kDa) if it is made in *R. temporaria* and *R. tigrina*. The antibody to the C-terminus should bind to the remaining fragment of Parapinopsin, which is the fragment that potentially could be produced after the site of mutation creating a stop codon.

Western Blot result analysis showed positive results with *X. laevis* for both N and C terminal antibodies that bound to a 38kDa protein, indicating Parapinopsin. The *X. laevis* Western Blot results were according to expectations as *X. laevis* was used as positive control. Wada *et al.*, in 2012 also found ~38kDa Parapinopsin protein in the parietal eye of green iguana through Western Blot analysis. *R. tigrina* Western Blot analysis results showed no protein band of 38kDa whereas a very clear Parapinopsin protein band showed for *X. laevis* on the same blot. Negative results were obtained with *R. tigrina* with both N and C terminal antibodies confirming the absence of Parapinopsin protein.

A point mutation at the same position was also found in *R. temporaria* Parapinopsin cDNA sequence like *R. tigrina*. But unlike *R. tigrina* gDNA the *R. temporaria* gDNA showed heterozygosity at the same position where a point mutation was found. The *R. temporaria* genome is very complicated in the sense of intraspecific genome size

variation and heterozygosity (C Matsuba and J Merila, 2006). In the current project, heterozygosity was also found during gDNA sequencing study of *R. temporaria* at the same position where a point mutation was found. *R. temporaria* has two copies of the gene for Parapinopsin one normal wild type and the other one is mutated. Western Blotting was repeated many times with *R. temporaria* protein samples. Every time negative results were obtained with both N and C terminal antibodies indicating the absence of Parapinopsin. For *R. temporaria* Western Blot studies, there is a need of more specific monoclonal antibodies (Tuttle *et al.*, 2013). Overall, Western Blot study revealed the detection of Parapinopsin protein by N and C terminal antibodies in *X.laevis* brain only.

6 Parapinopsin Study in Melanophores by Northern Blotting

Melanophores are photopigment cells expressed in *X.laevis* and melanin dispersion takes place in response to light. *X.laevis* dermal melanophores were cultured and their mRNA was used during the current study while *X.laevis* whole tadpole mRNA was used as positive control because Parapinopsin is expressed in *X.laevis* pineal gland.

6.1 Aim of Study

The aim of the current project was to study the expression of Parapinopsin mRNA in *X.laevis* melanophores by Northern Blotting.

6.2 Northern Blotting Steps

6.2.1 *X. laevis* Melanophore mRNA Extraction & Preparation of Denatured Agarose gel

X.laevis melanophores, whole *X.laevis* tadpoles (positive control) and kcl22 (Human leukaemia cell line, used as negative control) mRNA was extracted, quantified and analysed by 1% denatured agarose. The gel results are shown in Figure 6.1.

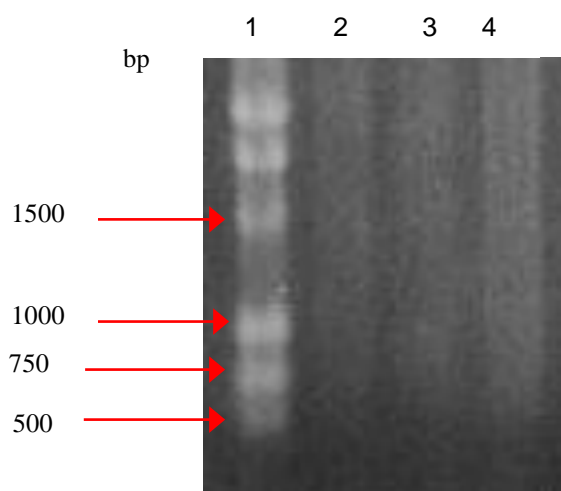


Figure 6.1 Electrophoretic Analysis of *X.laevis* melanophores, *X.laevis* tadpoles and Kcl22 mRNA

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2: *X.laevis* tadpole mRNA (positive control)

Lane 3: *X.laevis* melanophores mRNA

Lane 4: kcl22 mRNA (negative control)

A 1% denatured agarose gel was used to analyse the mRNA of *X.laevis* tadpole, *X.laevis* melanophore cell lines and kcl22 through electrophoresis in lanes 1, 2 and 3 respectively. The amount of mRNA loaded per lane is 6µg. The Kcl22 (human leukaemia cell line) mRNA was used as a negative control while *X.laevis* tadpole mRNA was used as positive control on the basis of absence and presence of Parapinopsin respectively.

6.2.2 Cross-Linking of mRNA to a Nylon Membrane

X.laevis melanophores, whole tadpoles and kcl22 (Human leukaemia cell line) mRNA once analysed by 1% denatured agarose gel, was transferred and cross linked to a nylon membrane.

6.2.3 DIG Labelled Single Stranded RNA Probe Synthesis

100-200ng of cloned 320bp Parapinopsin *X.laevis* cDNA PCR purified product shown in Figure 6.2, was used as a template by T7 RNA polymerase to synthesize DIG labelled single stranded RNA probe by *in vitro* transcription (DIG Northern starter kit, Roche). Promoter sequence for T7 RNA polymerase was added at 5' end of T7 forward primer.

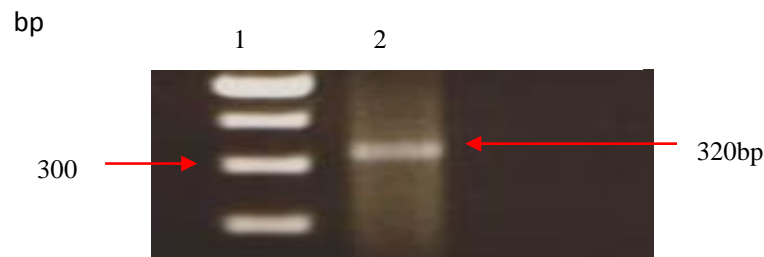


Figure 6.2 Electrophoretic Analysis of the PCR product of cloned pGEM PNP1

Lane 1: GeneRuler™ 100bp DNA Ladder

Lane 2: PCR product of PNP1 on cloned *X. laevis* cDNA

A 2% agarose gel was used to analyse the PNP1 PCR products through electrophoresis. In lane 2, the 320bp band indicates that the PCR reaction was successful and this gel purified product was used as template to synthesize DIG labelled single stranded RNA probe.

6.3 Northern Blot Results

The hybridization of cross linked *X.laevis* melanophores, Kcl22 and *X.laevis* mRNAs on a nylon membrane was carried out with DIG labelled RNA probe. Northern Blot results are shown below in Figure 6.3.

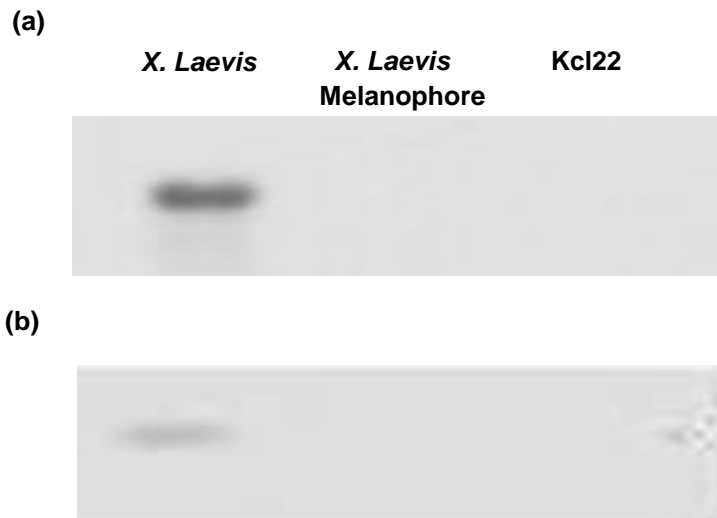


Figure 6.3 Northern Blot analyses for Parapinopsin in *X. laevis* tadpole, *X. laevis* Melanophore and Kcl22 mRNA with *X. laevis* Parapinopsin cDNA DIG labelled probe.

(a) The Figures illustrates the Northern Blot analysis results in *X. laevis* tadpoles (lane 1), *X. laevis* melanophore cell lines (lane 2) and negative control Kcl22 human leukaemia cell lines (lane 3) . Figure shows ~1kb band with *X.laevis* mRNA only. A repeat of whole experiment is shown in (b) with new mRNA samples.

The Northern Blot analysis results show the expression of Parapinopsin in *X. laevis* supporting the results of Western Blot studies. Parapinopsin is not expressed in *X. laevis* melanophore cell lines.

6.4 Summary of Northern Blotting Results

Sample detail	Expression of Parapinopsin
<i>X. laevis</i> Melanophore mRNA	No

<i>X. laevis</i> mRNA (positive control)	Yes
Kcl22 mRNA (negative control)	No

Table 6.1 Summary of Northern Blot Results

First column of the above table represents the sample details. *X.laevis* mRNA was used as a positive control during Northern Blotting. 2nd column shows the Northern Blot results suggesting the absence of Parapinopsin in *X.laevis* melanophores.

6.5 Discussion

Melanopsin is a non-visual opsin acting as a circadian photopigment which is expressed in the mammalian retina and involved in the biological clock. The same melanopsin is present in *X. laevis* dermal melanophores where its function is not fully known: it was discovered in 1998 in cultured cells. On exposure to light, the melanopsin in *X. laevis* melanophores is dispersed in melanin granules (Moraes *et al.*, 2014; Moraes *et al.*, 2013).

Photoreception takes place through the eyes and pineal, but melanophores are considered as photoreceptors. In sunlight the melanophores containing melanopsin become tan and in lower vertebrates help in camouflaging. The melanophores judge the seasonal light cycles and make a network of photo-responsive cells (Iyengar, 2013).

Yamashita *et al.*, in 2007 conducted RT-PCR analysis on the expression of Parapinopsin in lamprey pineal, brain and retina. They found the expression of Parapinopsin in lamprey pineal organ only. In amphibians, the expression of Parapinopsin is also reported in pineal gland (Koyanagi *et al.*, 2004).

During the current study it was aimed to check whether Parapinopsin is expressed in *X. laevis* melanophores or not, as melanophores are light sensitive cells. The Northern Blot study was conducted on *X. laevis* melanophore mRNA. The DIG labelled *X. laevis* cDNA was used as a probe. *X. laevis* tadpole mRNA was used as a positive control. The

Northern Blot study revealed the absence of expression of Parapinopsin in melanophores while positive results were obtained with *X. laevis* tadpoles mRNA.

Chapter 7 Expression of Parapinopsin Protein from *X. Tropicalis*

7.1 Aim of Study

This chapter describes the tailoring of a chemically synthesized codon optimised Parapinopsin gene from *X. tropicalis* (appendix 2) and its expression in GPCR plasmids *i.e.* pASK75-GPCR, pASK75-GPCR-GFP and pBAD33. The aim of the current studies is to overexpress the Parapinopsin protein for structural and functional studies using X-ray crystallography and high field NMR.

7.2 Tailoring of the Chemically Synthesized Parapinopsin Gene

The chemically synthesized cloned Parapinopsin gene in “pUC57 Kan” was tailored into the pASK75-GPCR plasmids by two step PCR.

7.2.1 Two Step PCR

The 110 nucleotides were added to the 5' end of the Parapinopsin gene by a two-step PCR reaction. For successful PCR the maximum length for the adaptor sequence is between 90-100nt. Hence total adaptor sequence is divided into almost two equal parts for a two-step PCR. The primer sequences are given in table 2.10 Material and the Method in Chapter 2.

The sequences for the *XbaI* restriction site, RBS (ribosomal binding site) and FLAG were added at the 5' end and *PstI*-His6-Stop-*HindIII* sequences were added at the 3' end for the subsequent ligation and expression of the Parapinopsin gene in the pASK-GPCR vectors. The 1% agarose gel PCR results are shown in Figure 7.1.

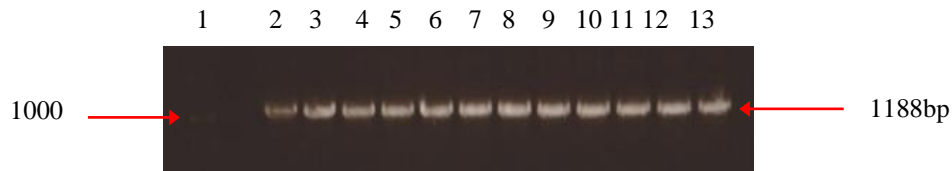


Figure 7.1 Electrophoretic Analysis of the PCR reaction using ParaP2 forward & common ParaP reverse primer on the chemical synthesized *X. tropicalis* Parapinopsin cDNA.

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2-13: PCR product of the Parap2 primer set

A 1% agarose gel was used to analyse Parap2 PCR products through electrophoresis. In lanes 2-13, the 1188bp band indicates that the PCR reaction was successful.

7.2.2 Restriction Digestion of Parapinopsin gene

The Parapinopsin PCR products were purified from agarose gel and digested by *HindIII*, *XbaI* and *PstI* restriction enzymes.

7.2.3 pASK75-GPCR Plasmids

Cells containing the pASK-GPCR plasmids (pASK-GPCR, pASK-GPCR-GFP) were streaked on LB agar plates, a single colony was grown in LB broth and after a miniprep the plasmids were digested with restriction enzymes *XbaI/HindIII* and *XbaI/PstI* respectively. The 1% agarose gel was used to analyse pASK-GPCR plasmids through electrophoresis and shown in Figure 7.2

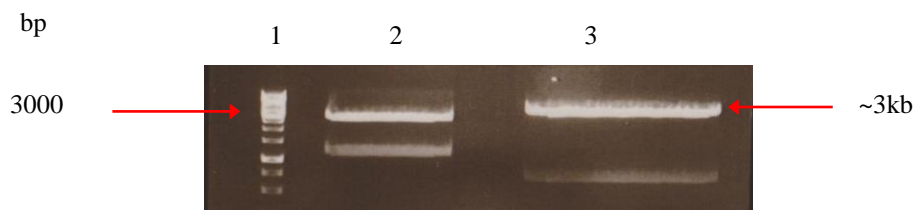


Figure 7.2 Electrophoretic Analysis of restriction digested pASK-GPCR and pASK-GPCR-GFP

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2: pASK-GPCR digested with restriction enzyme *XbaI/HindIII*

Lane 3: pASK-GPCR-GFP digested with restriction enzyme *XbaI/PstI*

A 1% agarose gel was used to analyse the pASK-GPCR plasmids by electrophoresis. In lane 2 and 3 the ~3kb band indicates that the reaction digestion reaction was successful.

The ~3kb fragment of pASK-GPCR and pASK-GPCR-GFP was purified from gel and used for the Parapinopsin cDNA ligation

7.2.4 Cloning of the Parapinopsin in the pASK75 plasmids

The gel purified Parapinopsin PCR products was cloned into the gel purified pASK-GPCR and pASK-GPCR-GFP plasmids. This was done by a ligation reaction as explained in section 2.9, Chapter 2. The reaction was analysed by PCR using primer set pASK75_3. A ~1000bp PCR product is expected. The results of PCR can be seen below in Figure 7.3.



Figure 7.3 Electrophoretic Analysis of ligation reaction PCR

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2: Ligation reaction ParaP PCR product of primer set pASK75_3 forward and reverse with pASK75-GPCR

Lane 3: Empty

Lane 4: Ligation reaction ParaP PCR product of primer set pASK75_3 forward and reverse with pASK75-GPCR-GFP

A 1% agarose gel was used to analyse PCR products of ligation mix through electrophoresis. In lane 2 and lane 4 the ~1000bp of pASK75_3 primer set PCR product bands indicate that the ParaP PCR products are successfully cloned in pASK75-GPCR and pASK75-GPCR-GFP vectors.

Half the amount of the ligation mixture was used to transform a 50µl aliquot of XL1Blue competent bacterial cells. The cells were then plated onto agar plates with appropriate antibiotics, and were left overnight in an incubator at 37°C. The 10 colonies on pASK75-GPCR and pASK75-GPCR-GFP plates were screened by means of PCR. The PCR was executed using a set of pASK75_3 forward and reverse primers. PCR products were analysed by electrophoresis, and the gel is illustrated in the Figure 7.4.

(a)



(b)



Figure 7.4 Electrophoretic Analysis of the PCR products of the pASK75-GPCR and pASK75-GPCR-GFP vectors Colony Screen

(a)

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 3, 5-9 & 11: Positive Colony screen PCR

Lanes 2, 4 & 10: Negative Colony screen PCR

The primer set pASK75_3 forward and reverse was used for the above colony screen and the expected PCR product size is 1014bp. Lanes 2, 4 and 10 have negative colony screen PCR products, but lanes Lane 3, 5-9 & 11 shown in red have band size of ~1000bp is present. These colonies were considered to be positive colonies. The whole colony was picked up from the agar plate and grown in 20ml of LB for mini prep, with the appropriate antibiotic.

(b)

Lane 1: GeneRuler™ 1kb DNA Ladder

Lane 2, 4 & 6: Positive Colony screen PCR

Lanes 1, 3, 5, 8, 9, 10 & 11: Negative Colony screen PCR

The primer set pASK75_3 forward and reverse was used for the above colony screen and the expected PCR product size is 1014bp. The lanes 1, 3, 5, 8, 9, 10 & 11 have negative colony screen PCR products, but lanes Lane 2, 4 & 6 shown in red have band size of ~1000bp is present. These colonies were considered to be positive colonies. The whole colony was picked up from the agar plate and grown in 20ml of LB for mini prep, with the appropriate antibiotics.

After the mini prep the cloned plasmids pASK75-GPCR and pASK75-GPCR-GFP, were analysed by restriction enzyme digests. The size of the fragment cut by the restriction enzymes should be ~1200bp. The results of restriction digestion are shown by 1% agarose gel electrophoresis, and are shown in Figure 7.5. Once positive colonies were confirmed by restriction digestion and were sent to Source Bioscience Life Sciences for sequencing.

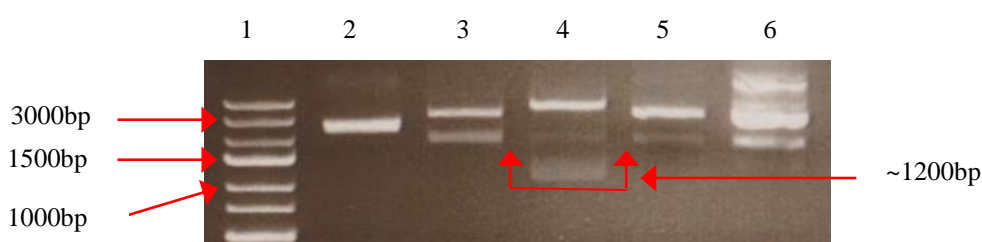


Figure 7.5 Electrophoretic Analysis of the PCR products of the pASK75-GPCR and pASK75-GPCR-GFP vectors Colony Screen

Lane 1: GeneRuler express DNA Ladder

Lane 2: Linearized Miniprep plasmid digested with restriction enzyme NdeI

Lanes 3: Double digestion of pASK-GPCR- miniprep plasmid with XbaI, HindIII restriction enzyme

Lanes 4: Double digestion of pASK-GPCR-GFP miniprep plasmid with XbaI, PstI restriction enzyme

Lanes 5: Uncut miniprep pASK-GPCR-GFP

Lanes 6: Uncut control plasmid pASK-GPCR-GFP

The 1% agarose gel was used to analyse the restriction digestion analysis. In lane 4 the fragment size of 1173bp represent the successful restriction digestion of pASK75-GPCR-GFP with XbaI & PstI restriction enzyme only. In lane 3 the restriction enzymes XbaI & HindIII failed to cut the ~1143bp fragment. The linearized plasmid digested by pASK-GPCR in lane 2 represents the successful ligation and recirculization of the plasmid.

Once the cloning of Parapinopsin in pASK-GPCR-GFP plasmid was confirmed, the miniprep plasmid was sent to Source Bioscience Life Sciences for sequencing.

7.2.5 Sequence Analysis of the pASK-GPCR-GFP

The sequencing so far has failed to generate successful results. The sequence reaction was repeated several times and so far time the results were not satisfactory. The work to get good a sequence readout is currently being pursued because the presence of even a single nucleotide mismatch will shift the whole frame and will produce hindrance during expression studies.

Chapter 8 Bioinformatics Based Study of Parapinopsin

Bioinformatics is used to conceptualize the macromolecules structure, function and evolutionary genetics by using computer based biological data. Bioinformatics studies based on statistical algorithms by using different online available softwares and tools information about Parapinopsin gene and protein.

8.1 Aim of Study

The aim of current study is to establish evolutionary relationship among divergent opsin proteins *i.e.* Parapinopsin, rhodopsin, pinopsin and Parietopsin, through biological multiple sequence alignment and phylogenetic tree by calculating and lining up the selected sequences best match to find out similarities and differences.

8.2 Protein Sequences used for Multiple Sequence Alignment

Human rhodopsin, *Xenopus tropicalis* cone opsin, all known Parapinopsin and non-visual opsin protein sequences to see homology, were used to align with *X. tropicalis* Parapinopsin protein sequences. All sequences were taken from NCBI (<http://www.ncbi.nlm.nih.gov/nuccore?term=parapinopsin>) and given with their accession numbers in appendix 5.

8.2.1 CLUSTAL W 2.1 Multiple Sequence Alignment

A computational based tool used to align multiple sequences called “CLUSTAL W” version 2.1 for “Multiple Sequence Alignment”, an important step to build a phylogenetic tree to link evolutionary points among different proteins sequences and provides information about substitutions over the period of time.

All known Parapinopsin along with non-visual and visual opsins protein sequences were aligned by using CLUSTAL W version 2.1 software, available at EBI (<http://www.ebi.ac.uk/Tools/msa/clustalw2>).

CLUSTAL 2.1 Multiple Sequence Alignment

[illegible]

Iguanapinopsin	NLLLT-----LRAVAAQQK-ESRTTTRAER-----EVTMRVIAMVLAFLVC--WLPYA	259
Humanrhodopsin	QLVFT-----VKEAAAQQQ-ESATTQKAER-----EVTMRVIIMVIAFLIC--WVPYA	269
tropconeopsin	HVWLT-----IRQVAQQQK-ESESTQKAER-----EVSRMVVVMIIAYIFC--WGPYT	284
ZebrafishPP	WLLYTLRQVLLEMVNNLGTFGMLFSAYYDLGHIFKHYKHCNAALTLVLFFVVEYHWLPYA	274
Cherry	-----KLILRQVSAMGCMEGGAAAKAEA-----KVARMVVLMVVTFLLIS--WLPYA	259
Cat	HLLHT-----LWQVAKLQVADSGSTAKVEV-----QVARMVVIMVMAFLLT--WLPYA	261
TroutPP	RILET-----LHQVSKLVLEGNSTTRVEI-----QVVRMVVVMVMAFLLS--WLPYA	262
XtropicalusPP	YLMWT-----LRQVAKLGVAEGGTTSKAEV-----QVSRMVIVMILAFLVC--WLPYA	261
LampreyPP	RLMWT-----LHQVAKLGMGESGSTAKAEA-----QVSRMVVVMVVAFLVC--WLPYA	254
Green	YLIWT-----LRQVAKVGLAQRGSTTKAEA-----QVSRMVIVMMAFLIC--WLPYA	257
IguanaPP	YLLWT-----LRQVAKFGVTQRGSTNKAEA-----QVSRMVIVMMAFLIC--WLPYA	257
iguanaparietopsin	NVIRS-----LHGLNKKVQLGGKSNPEEE-----FRAVVMVLVMVVAFMIC--WLPYT	259
Xparietopsin	SILRS-----LHQLNRKIEQQGGKTNPREE-----KRVMVIMVLMVLAFLIC--WLPYT	258
Zfparietopsin	NVLLT-----MRKINKSIECQGGKNCADN-----EHAVRMVLMAMIIAFFIC--WLPYT	255
	: : : : : *	
Iguanapinopsin	TFAMVVATNKDLVIQPALASLPYSFKTATVYNPIIYVFMNKQFRSCLLTYMRCGRPRD	319
Humanrhodopsin	SVAFYIFTHQGSNFGPIFMTIPAFFAKSAAIYNPVIYIMMKNQFRNCMLTICCGKNPLG	329
tropconeopsin	FFACFAAFNPGYNFHLAAAMPAYFAKSATIYNPIIYVFMNRQFRNCIYQLF--GKKVDD	342
ZebrafishPP	ALALTTVSKPEVQLAVLVKVLPIYMAKSSTVYNPLIYIYMKNQVSRTS-----	322
Cherry	TLALVVVSHPDAPISPLIGTPVYVLAKSSTVYNPLIYLYMKNQFRRYAVPFLLCGRDPWP	319
Cat	AFALTVIDSNIYINPVIGTIPAYLAKSSTVFNPPIIYIFMNRQFRDYALPCLLCGKNPWA	321
TroutPP	AFALSVIDPSLHINPLIATVPMYLAKSSTVYNPIIYVFMNRQFRDCAVPFLLCGLNPWA	322
XtropicalusPP	AFAMTVVANPGMHIDPIIATVPMYLTKTSTVYNPIIYIFMKNQFQECVIPFLFCGRNPWA	321
LampreyPP	LFAMIVVTKPDVYIDPVIATLPMYLTKTSTVYNPIIYIFMNRQFRDCAVPFLLCGRNPWA	314
Green	TFALVVVGNPQIYINPIIATIPMYMAKSSTFYNPPIIYIFMKNQFRDCLVRCLLCGRNPWA	317
IguanaPP	TFALVVVGNPQIYINPIIATVPMYMAKSSTFYNPPIIYIFMKNQFRDCLVRCLLCGRNPWA	317
iguanaparietopsin	VFALIVVFNPALNISPLAATIPTYLSKTSFVYNPIIYIFLNKQFRECAVEFITCGQVILT	319
Xparietopsin	VFALIVVINPQLYISPLAATLPTYFAKTSFVYNPIIYIFLNKQFRTYAVQCLTCGHINLD	318
Zfparietopsin	AISVLVVVNPEISIPPLIATMPMYFAKTSFVYNPIIYIFLTNKRRESSLEVLSGGRYISR	315
	: : : : : *	
Iguanapinopsin	ERETPPPRISVPRGPSSALEGSRNKVTPSASEGSGNDTRPS	360
Humanrhodopsin	DDEAS-----ATVSKTETSQVAPA-----	348
tropconeopsin	GSEVS-----STSRTVSSVSNSVSPA-----	365
ZebrafishPP	-----	
Cherry	SEEEGSEMQTTVATIN-----NKVSPS-----	341
Cat	AKEGRDSDTNTLTTT---VSKNTSVSPL-----	346
TroutPP	S-EPVGSEADTALSS---VSKNPRVSPQ-----	346
XtropicalusPP	AEKS-SSMETSISVTSGTPTKR-QQVAPA-----	348
LampreyPP	EPSSESATAASTSATSVTLASAPQVSPS-----	343
Green	SEQT-DEDDLEVSTIAPAPSSRRGKVAPV-----	345
IguanaPP	SDQA-EEDDLEVSTIAPAPSSKRKGVTPM-----	345
iguanaparietopsin	SQEE--DISVSAIPDEGKALCKINQVTPV-----	346
Xparietopsin	SLEE--DTESVSAQAENMLTPKTNQVAPA-----	345
Zfparietopsin	ETGG--PLMGSSMQRGQ-----SRVNPV-----	336

Table 8.1 Multiple Sequence Alignment CLUSTAL W

Above Table shows Multiple Sequence Alignment results of all known Parapinopsin, non-visual opsin, human rhodopsin and *X. tropicalis* cone opsin protein sequences i.e., Iguana Pinopsin, Human Rhodopsin, *X.tropicalis* cone opsin (tropconeopsin), Zebrafish Parapinopsin, Cherry salmon Parapinopsin, Cat fish Parapinopsin, Rainbow trout Parapinopsin, *X. tropicalis* Parapinopsin, Lamprey Parapinopsin, Green anole Parapinopsin, Common Iguana Parapinopsin, Iguana Parietopsin, *X. tropicalis* Parietopsin and Zebrafish Parietopsin. Interpretation of results is as follow:

- Abbreviated name of all protein sequence is given in 1st column
- All protein sequences are aligned as 60 residues/row
- * = aligned column with this symbol shows identical amino acids residues through all sequences
- : = aligned column with this symbol shows different but highly conserved amino acids

- . = aligned column with this symbol shows different amino acids that are somewhat similar
- Blank = aligned column with this symbol shows gaps

8.2.2 Multiple Sequence Alignment Result

SeqA	Name	Length	SeqB	Name	Length	Score
1	XtropicalusPP	348	2	LampreyPP	343	66.76
1	XtropicalusPP	348	3	Cat	346	56.94
1	XtropicalusPP	348	4	TroutPP	346	54.62
1	XtropicalusPP	348	5	ZebrafishPP	322	47.2
1	XtropicalusPP	348	6	Green anole	345	62.9
1	XtropicalusPP	348	7	Common Iguana	345	64.06
1	XtropicalusPP	348	8	Iguanapinopsin	360	43.39
1	XtropicalusPP	348	9	iguanaparietopsin	346	40.46
1	XtropicalusPP	348	10	tropconeopsin	365	38.51
1	XtropicalusPP	348	11	Humanrhodopsin	348	34.48
1	XtropicalusPP	348	12	Xparietopsin	345	40.58
1	XtropicalusPP	348	13	Zfparietopsin	336	36.31
1	XtropicalusPP	348	14	igparietopsin	346	40.46
1	XtropicalusPP	348	15	Cherry	341	52.49
2	LampreyPP	343	3	Cat	346	57.73
2	LampreyPP	343	4	TroutPP	346	58.6
2	LampreyPP	343	5	ZebrafishPP	322	47.83
2	LampreyPP	343	6	Green	345	58.89
2	LampreyPP	343	7	Common	345	60.64
2	LampreyPP	343	8	Iguanapinopsin	360	44.61
2	LampreyPP	343	9	iguanaparietopsin	346	39.36
2	LampreyPP	343	10	tropconeopsin	365	39.36

2	LampreyPP	343	11	Humanrhodopsin	348	34.4
2	LampreyPP	343	12	Xparietopsin	345	39.94
2	LampreyPP	343	13	Zfparietopsin	336	34.23
2	LampreyPP	343	14	igparietopsin	346	39.36
2	LampreyPP	343	15	Cherry	341	53.37
3	Cat	346	4	TroutPP	346	65.9
3	Cat	346	5	ZebrafishPP	322	47.83
3	Cat	346	6	Green	345	56.52
3	Cat	346	7	Common	345	57.1
3	Cat	346	8	Iguanapinopsin	360	41.91
3	Cat	346	9	iguanaparietopsin	346	37.28
3	Cat	346	10	tropconeopsin	365	36.99
3	Cat	346	11	Humanrhodopsin	348	36.42
3	Cat	346	12	Xparietopsin	345	40.0
3	Cat	346	13	Zfparietopsin	336	33.63
3	Cat	346	14	igparietopsin	346	37.28
3	Cat	346	15	Cherry	341	54.25
4	TroutPP	346	5	ZebrafishPP	322	48.45
4	TroutPP	346	6	Green	345	52.17
4	TroutPP	346	7	Common	345	53.04
4	TroutPP	346	8	Iguanapinopsin	360	42.49
4	TroutPP	346	9	iguanaparietopsin	346	38.44
4	TroutPP	346	10	tropconeopsin	365	33.82
4	TroutPP	346	11	Humanrhodopsin	348	34.97
4	TroutPP	346	12	Xparietopsin	345	41.74
4	TroutPP	346	13	Zfparietopsin	336	34.82
4	TroutPP	346	14	igparietopsin	346	38.44
4	TroutPP	346	15	Cherry	341	54.25

5	ZebrafishPP	322	6	Green	345	44.41
5	ZebrafishPP	322	7	Common	345	45.34
5	ZebrafishPP	322	8	Iguanapinopsin	360	36.34
5	ZebrafishPP	322	9	iguanaparietopsin	346	32.61
5	ZebrafishPP	322	10	tropconeopsin	365	33.85
5	ZebrafishPP	322	11	Humanrhodopsin	348	30.12
5	ZebrafishPP	322	12	Xparietopsin	345	36.02
5	ZebrafishPP	322	13	Zfparietopsin	336	32.92
5	ZebrafishPP	322	14	igparietopsin	346	32.61
5	ZebrafishPP	322	15	Cherry	341	53.42
6	Green	345	7	Common	345	93.04
6	Green	345	8	Iguanapinopsin	360	41.45
6	Green	345	9	iguanaparietopsin	346	39.42
6	Green	345	10	tropconeopsin	365	35.07
6	Green	345	11	Humanrhodopsin	348	35.07
6	Green	345	12	Xparietopsin	345	40.58
6	Green	345	13	Zfparietopsin	336	35.12
6	Green	345	14	igparietopsin	346	39.42
6	Green	345	15	Cherry	341	51.32
7	Common	345	8	Iguanapinopsin	360	41.45
7	Common	345	9	iguanaparietopsin	346	38.84
7	Common	345	10	tropconeopsin	365	34.78
7	Common	345	11	Humanrhodopsin	348	35.36
7	Common	345	12	Xparietopsin	345	41.16
7	Common	345	13	Zfparietopsin	336	35.42
7	Common	345	14	igparietopsin	346	38.84
7	Common	345	15	Cherry	341	51.32
8	Iguanapinopsin	360	9	iguanaparietopsin	346	37.28

8	Iguanapinopsin	360	10	tropconeopsin	365	41.67
8	Iguanapinopsin	360	11	Humanrhodopsin	348	42.82
8	Iguanapinopsin	360	12	Xparietopsin	345	39.71
8	Iguanapinopsin	360	13	Zfparietopsin	336	35.71
8	Iguanapinopsin	360	14	igparietopsin	346	37.28
8	Iguanapinopsin	360	15	Cherry	341	39.59
9	iguanaparietopsin	346	10	tropconeopsin	365	31.79
9	iguanaparietopsin	346	11	Humanrhodopsin	348	33.24
9	iguanaparietopsin	346	12	Xparietopsin	345	73.33
9	iguanaparietopsin	346	13	Zfparietopsin	336	51.49
9	iguanaparietopsin	346	14	igparietopsin	346	100.0
9	iguanaparietopsin	346	15	Cherry	341	37.83
10	tropconeopsin	365	11	Humanrhodopsin	348	37.36
10	tropconeopsin	365	12	Xparietopsin	345	34.2
10	tropconeopsin	365	13	Zfparietopsin	336	29.46
10	tropconeopsin	365	14	igparietopsin	346	31.79
10	tropconeopsin	365	15	Cherry	341	36.66
11	Humanrhodopsin	348	12	Xparietopsin	345	33.62
11	Humanrhodopsin	348	13	Zfparietopsin	336	28.27
11	Humanrhodopsin	348	14	igparietopsin	346	33.24
11	Humanrhodopsin	348	15	Cherry	341	34.31
12	Xparietopsin	345	13	Zfparietopsin	336	52.08
12	Xparietopsin	345	14	igparietopsin	346	73.33
12	Xparietopsin	345	15	Cherry	341	39.59
13	Zfparietopsin	336	14	igparietopsin	346	51.49
13	Zfparietopsin	336	15	Cherry	341	33.33
14	igparietopsin	346	15	Cherry	341	37.83

Table 8.2 Multiple Sequence Alignment Result (MSA) Summary

(<http://www.ebi.ac.uk/Tools/msa/clustalw2>)

Above table explains the pairwise alignment of “seq A” in column 1 and “seq B” in column 4.

Column 2 and 5 explains the name of protein sequence, column 3 and 6 explains the total length of protein sequence and column 7 explains the score of aligned sequence.

Maximum sequence similarity scored by *X. tropicalis* Parapinopsin and Lamprey Parapinopsin which is 66.76%. The sequence similarity shared between *X. tropicalis* Parapinopsin and Zebrafish Parapinopsin is 47.2%. Sequence similarity between *X. tropicalis* Parapinopsin and Human Rhodopsin protein is 34.48% and *X. tropicalis* Parapinopsin sequence is 40.58% similar with *X. tropicalis* Parietopsin.

8.3 Parapinopsin Protein Phylogenetic Tree

All protein sequences mentioned in section 8.2 were used to build a Phylogenetic tree which determines the evolutionary relations among species.

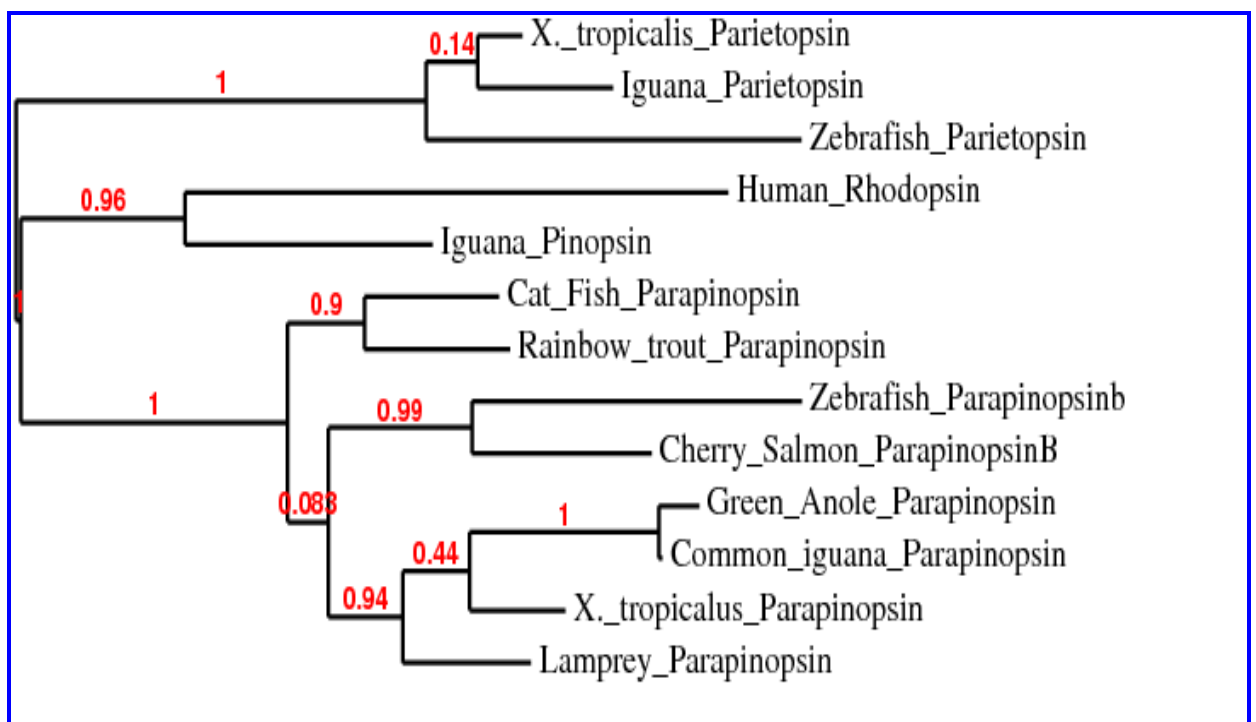


Figure 8.2 Parapinopsin phylogenetic tree with other opsins

(http://www.phylogeny.fr/version2_cgi/simple_phylogeny.cgi)

Figure above shows the Parapinopsin phylogenetic tree constructed by “TreeDyn 198.3” program on the basis of conserved amino acid alignment. Each branch node has bootstrap values (higher bootstrap value indicates strong support for a clade). The name of each Parapinopsin/Opsin protein is given at the right side of each cluster. All the sequences were taken from NCBI Genbank databases.

8.3 Parapinopsin DNA Sequences used for Multiple Sequence Alignment

All known Parapinopsin including newly found *X. laevis*, *R. temporaria* and *R. tigrina* Parapinopsin and non-visual opsin DNA sequences were used to align. All sequences were taken from NCBI (<http://www.ncbi.nlm.nih.gov/nuccore?term=parapinopsin>) and given with their accession numbers in appendix 6.

8.3.1 Multiple sequence comparison by log-expectation (MUSCLE)

To find sequence similarity among different Parapinopsin DNA sequences a computer based tool used to align multiple DNA sequences called “Multiple sequence comparison by log-expectation (MUSCLE) by Robert C. Edgar Version 3.5.

	10	20	30	40	50	60
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----	-----	-----	ATGGACTTTT	-----	TC
Cherry_salmon_p	-----	GGGGGAGTGTGGTGTGGAGGTTCTTACCATCT	-----	-----	-----	CC
Green_anole_pp	-----	-----	-----	-----	ATGGATTC	TC
common_Iguana	-----	-----	TTTCATAGGCATCTAAAGAA	-----	AAGATGAC	TC
Lamprey_pp	-----	GTGAGTGTATTTTATAAC	-----	TCGGTGCTTGA	-----	TT
R._temporaria_p	-----	-----	TGTTGCCCTGACAATTTTCGGTGACTAAAAAGTC	-----	-----	TC
R._tigrina_pp	-----	-----	TGTTGCCCTGACAATTTTCGGTGACTAAAAAGTC	-----	-----	TC
X.tropicalis_pp	-----	-----	TGTTGCCCTGACAATTTTCGGTGACTAAAAAGTC	-----	-----	TC
X._laevis_pp	-----	-----	TGTTGCCCTGACAATTTTCGGTGACTAAAAAGTC	-----	-----	TC
Rainbow_trout_p	ATCATAACACAGGTGTCTGGCCAGGTGGGTAGTGCTCACCCTC	-----	AACTCCACTGCC	-----	-----	TC
Cat_Fish_pp	-----	ATGGCATCCATTATTCTAATCAACTT	-----	-----	-----	TT
	70	80	90	100	110	120
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	CCAGAAATTCCTCAACGA	-----	-----	-----	-----	-----
Cherry_salmon_p	AAGAGAGTGAGAGCCTGTGTGGTGTGTGCTTGTGTTGTTATATATCCATGATGCAGCTG	-----	-----	-----	-----	-----
Green_anole_pp	TCTGGAC	-----	-----	-----	-----	-----
common_Iguana	TGAAGACTGACAGACAGCATAAAACAACAAAGCTTGAGCAGTATAGGCATAAAAAAAGA	-----	-----	-----	-----	-----
Lamprey_pp	TGAAGATCGTCTGTTTCG	-----	-----	-----	-----	TCAA
R._temporaria_p	TCAGAACCAAATGGCCG	-----	-----	-----	-----	ACGA
R._tigrina_pp	TCAGAACCAAATGGCCG	-----	-----	-----	-----	ACGA

X.tropicalis_pp	TCAGAACCAAATGGCCG-----ACGA
X._laevis_pp	TCAGAACCAAATGGCCG-----ACGA
Rainbow_trout_p	CCAAAAAGAACAGAATG-----GACCACCA
Cat_Fish_pp	CAGAGACTGACACACTG-----

	130	140	150	160	170	180
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	--CTCTTCCACT----	CAC	TATGATGGCCCTGGT			TTG-----
Cherry_salmon_p	CCAGCCTCTCTC----	CCA	ACGCTTCCTCCTAC			CTGGGACCAT
Green_anole_pp	-----	ACCA	ACACTT--TGTCAC			CTAATGCAAG
common_Iguana	GTATTTTATTTTGGTATC	CTA	ACAACAAATGGATTTT			TGGACACCAGCACTTTGTCAAT
Lamprey_pp	G-----TAAG----	AGC	CATGGAGAACTTGAC			CTCGCTCGAC
R._temporaria_p	GGCTCTTCTCCC----	ACCA	ATGATGAATGTGAC			CAA---TGAA
R._tigrina_pp	GGCTCTTCTCCC----	ACCA	ATGATGAATGTGAC			CAA---TGAA
X.tropicalis_pp	GGCTCTTCTCCC----	ACCA	ATGATGAATGTGAC			CAA---TGAA
X._laevis_pp	GGCTCTTCTCCC----	ACCA	ATGATGAATGTGAC			CAA---TGAA
Rainbow_trout_p	ACAGCTTCTCCCCAATCT	CCA	TGGAAACATCTCCT			CCTCCCCGGG
Cat_Fish_pp	---CATCTCGG----	CTC	AGTAAATGATCACA			

	190	200	210	220	230	240
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	-----	AAGGTGC	---	CTCTTCCCCGGGCAGGATTCATCAGCTGTC		
Cherry_salmon_p	-----	CCCCGGAGGGCAAGGAGC	---	TCTTCCCCGAGCCGGCTTCATCACCTCTC		
Green_anole_pp	CACA-----	GTACGAGTGGTCT	---	TGATGC-----	TTGGTTATACTATAATAGC	
common_Iguana	GAATGCAACCACAGTACGAGTGGTCT	---	TGATGCCCCGTAT			TGGTTATACTATAATAGC
Lamprey_pp	CTCCTGCCCAACG----	GCGAGGTCCCGT		TGATGCCCCGCTACGGCTTCACCATCCTCGC		
R._temporaria_p	GAGATGCACCCTG----	GGAAGGTTC	---	TGATGCCCCGGATTGGTTACACCATTCTGGC		
R._tigrina_pp	GAGATGCACCCTG----	GGAAGGTTC	---	TGATGCCCCGGATTGGTTACACCATTCTGGC		
X.tropicalis_pp	GAGATGCACCCTG----	GGAAGGTTC	---	TGATGCCCCGGATTGGTTACACCATTCTGGC		
X._laevis_pp	GAGATGCACCCTG----	GGAAGGTTC	---	TGATGCCCCGGATTGGTTACACCATTCTGGC		
Rainbow_trout_p	-----	CTCGGTCAGTGAAGCGC	---	TGCTCTCCAGGACGGGCTTCACCATCCTGGC		
Cat_Fish_pp	-----			TCATGCCAAGAATAGGCTACACCATTTTGTC		

	250	260	270	280	290	300
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	CCTCCTCATGGCCGTGTTCTCCATC			ATCTGTCGTGTGAACGCTACAGTGATAGTCGT		

Cherry_salmon_p	CATCGTCATGGCTATGTTACCGTACCGGCATTGTGTTGAACTCCACAGTGATCGTGGT
Green_anole_pp	AATTATTATGGCAACATCTTGCACCTTTATCAGTAATTTTGAATACAGCCGTCATTGCCAT
common_Iguana	AATTATTATGGCAACATCTTGCACCTTTATCAGTAATTTTGAATACAACGTGTCATTGCAGT
Lamprey_pp	CGTGATCATGGCCGTGTTACCATCGCCTCGCTCGTGCTCAACAGCACCGTCGTCATCGT
R._temporaria_p	TTTAATAATGGCCGTATTCTGTGCCGCCGCTCTCTTTCTCAACGTCACGGTGATTGTGGT
R._tigrina_pp	TTTAATAATGGCCGTATTCTGTGCCGCCGCTCTCTTTCTCAACGTCACGGTGATTGTGGT
X.tropicalis_pp	TTTAATAATGGCCGTATTCTGTGCCGCCGCTCTCTTTCTCAACGTCACGGTGATTGTGGT
X._laevis_pp	TTTAATAATGGCCGTATTCTGTGCCGCCGCTCTCTTTCTCAACGTCACGGTGATTGTGGT
Rainbow_trout_p	GGTGATCATGGTGTGTTCTCGGTGTCCGGAGTCTGTATGAACGTCCTGGTCATCATGGT
Cat_Fish_pp	AATAATTATGGCTTTGTCTCCACTTTTGCATCATCTTGAATATGGTGGTGATTATTGT
	310 320 330 340 350 360
	=====+=====+=====+=====+=====+=====+
Zebrafish_pp_b	CACCCGTCCGGCACAAACAGCTGCGGCAGCCGCTCAATTTTGCCTGGTTAACCTCGCCGT
Cherry_salmon_p	GTGCTGATGAACAAGCAGCTGAGGCAACCTCTGAACCTACCCCTGGTCAACATGGCCGT
Green_anole_pp	AACTATTAACTACAGGCAGCTTCGTGAGCCCATCAATTATCTCTGGTTAATTTAGCAAT
common_Iguana	AACTATAAACTACAGACAGCTTCGGCAGCCCATCAATTATCTCCTGGTTAATTTAGCTAT
Lamprey_pp	CACCCGTCCGCCACCGCCAGCTACGCCACCGCTCAACTTCTGCTCGTCAACCTCGCCGT
R._temporaria_p	GACTTTCAAATATCGCCAACTGCGCCACCCAATCAACTACTCGCTGGTCAACCTGGCCAT
R._tigrina_pp	GACTTTCAAATATCGCCAACTGCGCCACCCAATCAACTACTCGCTGGTCAACCTGGCCAT
X.tropicalis_pp	GACTTTCAAATATCGCCAACTGCGCCACCCAATCAACTACTCGCTGGTCAACCTGGCCAT
X._laevis_pp	GACTTTCAAATATCGCCAACTGCGCCACCCAATCAACTACTCGCTGGTCAACCTGGCCAT
Rainbow_trout_p	GACGATGAGACACCGAAAGCTGCGCCAGCCTTTGAACTACGCCCTAGTCAACCTGGCAGT
Cat_Fish_pp	GACAGTCCGGTACAAACAACCTGCGTTCAGCCACTAAACTATGCTCTGGTTAATCTAGCAGT
	370 380 390 400 410 420
	=====+=====+=====+=====+=====+=====+
Zebrafish_pp_b	GGCCGACCTGGGCACCACTAACAGGAACCGTGCCATCTGTGGTGACCAATGCTGTAGG
Cherry_salmon_p	AGCTGACCTGGGACAGCGCTGACTGGTGGGGTTCTGTCTGTGGTCAACAAATGCTCTGGG
Green_anole_pp	AGCAGACCTTGGAGCAGCTTTGTTAGGAGGAAGCCTGAATGTAGAAACAAATGCCGTTGG
common_Iguana	AGCAGACCTTGGAGCAGCTTTGTTAGGAGGAAGTCTGAACGTAGAAACAAATGCTGTTGG
Lamprey_pp	GGCCGACCTGGGCGTCACCGTGTTCGGCCAGCCTCGTCTGGAGACCAACGCCGTCG-
R._temporaria_p	CGCCGACCTTGGCGTCACCGTGTCTGGCGGGGCCTTGACAGTGGAGACAAATGCCGTGGG
R._tigrina_pp	CGCCGACCTTGGCGTCACCGTGTCTGGCGGGGCCTTGACAGTGGAGACAAATGCCGTGGG
X.tropicalis_pp	CGCCGACCTTGGCGTCACCGTGTCTGGCGGGGCCTTGACAGTGGAGACAAATGCCGTGGG
X._laevis_pp	CGCCGACCTTGGCGTCACCGTGTCTGGCGGGGCCTTGACAGTGGAGACAAATGCCGTGGG
Rainbow_trout_p	GGCTGACCTGGGCTGTGCTCTGTGTTGGGGGACTGCCACGATGGTCACCAACGCTATGGG
Cat_Fish_pp	GGCTGATCTTGGCTGTCTGTGTTTGGGGGCTACTCACAGCAGTCACCAATGCCATGGG
	430 440 450 460 470 480

	=====+=====+=====+=====+=====+=====+=====+
Zebrafish_pp_b	CTACTACATTATGGGACCGATCGGCTGTGTATTGGAGGGATTTTGTGTTGCATTTTTTTGG
Cherry_salmon_p	GTACTTCTCCCTGGGACCGACCGGCTGTATCATAGAGGGATTCTCTGTGGCTCTGTTTGG
Green_anole_pp	ATATTACAACCTGGGCAGAGTTGGCTGTGTACCGAAGGTTTTGCTATGGCATTTTTTTGG
common_Iguana	GTATTACAACCTGGGCAGAGTTGGCTGTGTACCGAAGGTTTTGCTATGGCATTTTTTTGG
Lamprey_pp	-TACTTCAACCTCGGCCCGCTCGGCTGCGTCATCGAAGGTTTCGCCGTGCTTTCTTCGG
R._temporaria_p	CTACTTTAACCTGGGCAGAGTGGGCTGCGTCATTGAGGGTTTCGCCGTGGCATTTTTTTGG
R._tigrina_pp	CTACTTTAACCTGGGCAGAGTGGGCTGCGTCATTGAGGGTTTCGCCGTGGCATTTTTTTGG
X.tropicalis_pp	CTACTTTAACCTGGGCAGAGTGGGCTGCGTCATTGAGGGTTTCGCCGTGGCATTTTTTTGG
X._laevis_pp	CTACTTTAACCTGGGCAGAGTGGGCTGCGTCATTGAGGGTTTCGCCGTGGCATTTTTTTGG
Rainbow_trout_p	CTACTTCAGCATGGGGCGCTAGGCTGCGTGCTGGAGGGTTTCGCTGTGGCTTTCTTTGG
Cat_Fish_pp	CTACTTCAGTCTTGGCAGGCTAGGATGTGTTTGGAAGGATTTCCTGTTGCCTTCTTTGG
	490 500 510 520 530 540
	=====+=====+=====+=====+=====+=====+=====+
Zebrafish_pp_b	TATCTCCGGCTGTGCACCGTGGCTCTGATCGCAGTGGACCGTCTGTTTGTGGTGTGCAG
Cherry_salmon_p	TATCAGCTCACTGTGCACGGTAGCTTTGATGGCATAAGAGGATGTTTGTGGTGTGCAG
Green_anole_pp	TATTGTGGCTTTGTGCACCATAGCAGTGATAGCTGTTGACCGCGCAATTGTAATCGCCAA
common_Iguana	CATCGTGGCTTTGTGCACCATAGCAGTAATAGCTGTGACCGTGCAATTGTGATTGCCAA
Lamprey_pp	CATCGCCGCTCTGTGCACGATCGCCGTGATCGCCGTGATCGCTTCGTGGTGGTGTGCAG
R._temporaria_p	CATCGCGGCTTTGTGCACCATTGCAGTAATAGCGTTGGACCGAGTGTTCTGGTGTGCAG
R._tigrina_pp	CATCGCGGCTTTGTGCACCATTGCAGTAATAGCGTTGGACCGAGTGTTCTGGTGTGCAG
X.tropicalis_pp	CATCGCGGCTTTGTGCACCATTGCAGTAATAGCGTTGGACCGAGTGTTCTGGTGTGCAG
X._laevis_pp	CATCGCGGCTTTGTGCACCATTGCAGTAATAGCGTTGGACCGAGTGTTCTGGTGTGCAG
Rainbow_trout_p	TATTGCAGCTCTGTGTTCACTGGCCGTAATAGCTGTGGACCGTTATGTGTGGTGTGTGCG
Cat_Fish_pp	TATTGCTGCTTTGTGCTCGGTGCGAGTTATTGCTGTTGACCGCTACATGTTAGTATGCAG
	550 560 570 580 590 600
	=====+=====+=====+=====+=====+=====+=====+
Zebrafish_pp_b	ACCTCTGGGCTCCATCAGCTTTCAGTGCAACACGCCGAGGAGGCCTGTTGTCCTGCTG
Cherry_salmon_p	GCCGCTAGGACCAATCTCTTCCAGACCAAGCATGCGGTTGGGGTGTGGCCTTGCTATG
Green_anole_pp	GCCGATGGGAACAATTACATTTACTACCAGAAAGCTATGATTGGTGTAGCTGTGTCCTG
common_Iguana	GCCTATGGGAACACTAATTTTACCACCAGGAAAGCAATGATTGGAGTCGCTATCTCCTG
Lamprey_pp	GCCGCTGGGCACGCTGATGTTACGCGGCGCCACGCGCTGCTGGGCATCGCTGGGCGCTG
R._temporaria_p	GCCAATGGGCACACTCACAATTCACCCCCAAACAAGCACTGGCCGGCATTGCTGCTTCCTG
R._tigrina_pp	GCCAATGGGCACACTCACAATTCACCCCCAAACAAGCACTGGCCGGCATTGCTGCTTCCTG
X.tropicalis_pp	GCCAATGGGCACACTCACAATTCACCCCCAAACAAGCACTGGCCGGCATTGCTGCTTCCTG
X._laevis_pp	GCCAATGGGCACACTCACAATTCACCCCCAAACAAGCACTGGCCGGCATTGCTGCTTCCTG
Rainbow_trout_p	ACCAATGGGGGGCGTCTGTTCCAGACCAAGGCACGCACTCGGAGGGGTGCTCCTGTCTCTG
Cat_Fish_pp	ACCATAGGTGCACTGATGTTCCAAACCAACATGCCCTGGCAGGCGTGGTTTTTTCTCTG

	610	620	630	640	650	660
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	GCTTTGGTCTCTGATCTGGAACACTCCGCCTCTGCTGGGCTGGGGAAGCTACAGCTGGA					
Cherry_salmon_p	GGTTTGGTCTCTCCTCTGGAACACCCGCCCTATTCCGGCTGGGGAAGGTACGAGCTGGA					
Green_anole_pp	GATATGGTCTCTTGTGTTGGAACACACCTCCTCTTTTTGGTTGGGGAGGATACCAGATGGA					
common_Iguana	GATATGGTCTCTTGTGTTGGAACACGCCCTCCTCTTTTTGGTTGGGGAGGATACCAGATGGA					
Lamprey_pp	GCTCTGGTCGTTCTGTTGGAACACGCCGCCGCTCTTCGGCTGGGGCAGCTACGAGCTGGA					
R._temporaria_p	GATCTAGTCGCTCATATGGAATACGCCCCCGCTGTTCCGGCTGGGGTAGCTACGAGCTGGA					
R._tigrina_pp	GATCTAGTCGCTCATATGGAATACGCCCCCGCTGTTCCGGCTGGGGTAGCTACGAGCTGGA					
X.tropicalis_pp	GATCTGGTCGCTCATATGGAATACGCCCCCGCTGTTCCGGCTGGGGTAGCTACGAGCTGGA					
X._laevis_pp	GATCTGGTCGCTCATATGGAATACGCCCCCGCTGTTCCGGCTGGGGTAGCTACGAGCTGGA					
Rainbow_trout_p	GGTCTGGTCGTTTTTATGGAACACCCGCCCTCTGTTTGGCTGGGGCAGCTTTCAACTGGA					
Cat_Fish_pp	GGTCTGGTCCTTCATATGGAACACGCCACCACTGTTTGGATGGGGCAGTTACCAGCTGGA					
	=====+=====+=====+=====+=====+=====+					
	670	680	690	700	710	720
Zebrafish_pp_b	GGGAGCGGGCAGCTCCTGCGGGCCGCACTGGCAGAGTCGAGAACTCAGAGACGTGTCTTA					
Cherry_salmon_p	GGGGGTGGGACGTCCTGTGCCCCGGAAGTGGCACAACCGGACCCCAATAATGTGTCTATA					
Green_anole_pp	AGGCGTTATGACCTCCTGCGCTCCAGATTGGGCTAACAGTGATCCTATCAATGTTTCTCTA					
common_Iguana	AGGCGTCATGACCTCCTGCGCTCCAGATTGGTACAACAGTGATCCTATTAATGTCTCTCTA					
Lamprey_pp	GGGCGTCGGACGTCCTGCGCGCCCCACTGGTACAGCGCGACCCCGCAACGTGTCTGTA					
R._temporaria_p	AGGGGTGATGACATCCTGCGCCCCCAACTGGTACAGCGCGACCCAGTCAACATGTCTCTA					
R._tigrina_pp	AGGGGTGATGACATCCTGCGCCCCCAACTGGTACAGCGCGACCCAGTCAACATGTCTCTA					
X.tropicalis_pp	AGGGGTGATGACATCCTGCGCCCCCAACTGGTACAGCGCGACCCAGTCAACATGTCTCTA					
X._laevis_pp	AGGGGTGATGACATCCTGCGCCCCCAACTGGTACAGCGCGACCCAGTCAACATGTCTCTA					
Rainbow_trout_p	GGGGGTACGACCTCCTGCTCCCCAACTGGTACAGCAGGAAACCAGGAAACATGTCTCTA					
Cat_Fish_pp	GGGTGTGATGACCTCCTGTGCTCCCAACTGGTACAGGAGAGACCCAGTCAATGTCTCTCTA					
	=====+=====+=====+=====+=====+=====+					
	730	740	750	760	770	780
Zebrafish_pp_b	CATCATCTGCTACTTCTCCGCTCTGCTTCGCGGTGCCATTCCGCATCATCTGGTGTCTGTA					
Cherry_salmon_p	TATCCTGTGCTACTTCTCTGCTGTGTTTTGCGGTGCCCTTCCTCATCATCGTGGCCTCTTA					
Green_anole_pp	CATTATCTGCTACTTTCTCTTCTGCTTTACAATCCCATTCATCACAATCTTGGCCTCCTA					
common_Iguana	CATTGTCTGCTACTTTCTATTCTGCTTTACGATTCCCTTCCTCACTATTTGGTCTCTCTA					
Lamprey_pp	CATTACGAGCTACTTCGCTTCTGCTTCGCCATCCCCCTTCCTCGTCATCGTGGTGGCGTA					
R._temporaria_p	CATCGTCTGCTACTTCTCCTTCTGCTTTGCCATCCCATTCCTCATCATCGTGGGGTCTCTA					
R._tigrina_pp	CATCGTCTGCTACTTCTCCTTCTGCTTTGCCATCCCATTCCTCATCATCGTGGGGTCTCTA					
X.tropicalis_pp	CATCGTCTGCTACTTCTCCTTCTGCTTTGCCATCCCATTCCTCATCATCGTGGGGTCTCTA					
X._laevis_pp	CATCGTCTGCTACTTCTCCTTCTGCTTTGCCATCCCATTCCTCATCATCGTGGGGTCTCTA					
Rainbow_trout_p	CATCATCTTTACTTCTCTGCTCTGCTTTGCCATTCTTTCTCCATCATCATGGTCTCTCTA					
Cat_Fish_pp	TATCCTGTGCTACTTCATGCTCTGTTTTGCGCTGCCCTTTTGCCACCATTATCTTCTCTTA					

	790	800	810	820	830	840
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	TTCA	TGGCTGCTCTACACGCTGAGACAAGTTT	---	ACTGCAAATCGTAAACAATCTGGG		
Cherry_salmon_p	CTCTAA	ACTCATTACACCTTAAGACAGGCTCTCTGCAATGGGCTGTATGGA	----	AGG		
Green_anole_pp	CGGT	TACCTGATATGGACTCTGCGTCAGGTAGCTAAAGTTGGACTAGCACA	----	GCG		
common_Iguana	TGGCT	TACCTGCTATGGACTCTGCGCCAGGTTGCTAAATTTGGAGTAACACA	----	GCG		
Lamprey_pp	CGGCCG	CCTCATGTGGACCTCCACCAGGTGGCCAAGCTGGGGATGGGCGA	----	GAG		
R._temporaria_p	CGGGT	TACCTGATGTGGACTCTGCGGCAGGTTGCCAAGTTGGGTGTAGCGGA	----	AGG		
R._tigrina_pp	CGGGT	TACCTGATGTGGACTCTGCGGCAGGTTGCCAAGTTGGGTGTAGCGGA	----	AGG		
X.tropicalis_pp	CGGGT	TACCTGATGTGGACTCTGCGGCAGGTTGCCAAGTTGGGTGTAGCGGA	----	AGG		
X._laevis_pp	CGGGT	TACCTGATGTGGACTCTGCGGCAGGTTGCCAAGTTGGGTGTAGCGGA	----	AGG		
Rainbow_trout_p	CGCTCG	CATCCTCTTCACACTGCACCAGGTGTCCAAGCTCAAGGTGTTTGA	----	GGG		
Cat_Fish_pp	CATGCAC	CTCCTTCATACCTTATGGCAGGTGGCTAAGTTGCAAGTGGCAGA	----	CAG		
		850	860	870	880	
890	900					
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	AACTTTT	GCGATGCTCTTTTCAGCATACTACGATTTGGGACATATTTTCAAACTATAA				
Cherry_salmon_p	G----	GGTG-----CAGCAGCCAAAGCGGAGGCTAAGGTGGCC	-----			
Green_anole_pp	A----	GGCT-----CAACAACATAAGCGGAAGCACAAAGTGTCA	-----			
common_Iguana	A----	GGTT-----CAACAACATAAGCGGAAGCTCAGGTGTCA	-----			
Lamprey_pp	C----	GGCA-----GCACCGCCAAAGCGGAGGCCAGGTGTCTG	-----			
R._temporaria_p	A----	GGTA-----CAACGAGTAAAGCCGAGGTGCAAGTCTCC	-----			
R._tigrina_pp	A----	GGTA-----CAACGAGTAAAGCCGAGGTGCAAGTCTCC	-----			
X.tropicalis_pp	A----	GGTA-----CAACGAGTAAAGCCGAGGTGCAAGTCTCC	-----			
X._laevis_pp	A----	GGTA-----CAACGAGTAAAGCCGAGGTGCAAGTCTCC	-----			
Rainbow_trout_p	C----	AACA-----GCACCACC	CGGTGGAGATACAGGTTGTG	-----		
Cat_Fish_pp	T----	GGAA-----GCACGGCCAAAGTTGAGGTTCAAGTGGCC	-----			
	910	920	930	940	950	960
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	GCATTG	CAACGCAGCTTTAAGCTTAGTTTATTTTGGTTGTTGAATATCACTGGTTGCC				
Cherry_salmon_p	----	CGCATGGTGGTTCTCATGGTGGTTACTTTTCTGAT	-----	CAGCTGGCTGCC		
Green_anole_pp	----	CGCATGGTGAATTGTGATGGTTATGGCATTCTTGAT	-----	CTGCTGGCTTCC		
common_Iguana	----	CGCATGGTGAATGTAATGGTTATGGCATTCTTGAT	-----	CTGCTGGCTTCC		
Lamprey_pp	----	CGCATGGTGGTGGTCATGGTGGTGGCCTTCTCGT	-----	CTGCTGGCTGCC		
R._temporaria_p	----	CGTATGGTGAATTGTCATGATCTTGGCTTTCTGCT	-----	CTGTTGGCTCCC		
R._tigrina_pp	----	CGTATGGTGAATTGTCATGATCTTGGCTTTCTGCT	-----	CTGTTGGCTCCC		
X.tropicalis_pp	----	CGTATGGTGAATTGTCATGATCTTGGCTTTCTGCT	-----	CTGTTGGCTCCC		
X._laevis_pp	----	CGTATGGTGAATTGTCATGATCTTGGCTTTCTGCT	-----	CTGTTGGCTCCC		
Rainbow_trout_p	----	CGTATGGTGGTAGTCATGGTGATGGCCTTCTGCT	-----	CAGCTGGTTGCC		
Cat_Fish_pp	----	CGTATGGTGGTCATCATGGTCATGGCCTTTTGGCT	-----	TACATGGCTGCC		

	970	980	990	1000	1010	1020
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	TTACGCTGCTCTCGCTCTA	ACCGTCGTCTCCAAACCTG	AGGTCCAGCTGGCCGTGCT	GGT		
Cherry_salmon_p	CTACGCTACCTGGCCCTT	GTGTGGTGGTCTCCACCT	GATGCCCCATTAGCCCT	CTGCT		
Green_anole_pp	TTACGCAACTTTTGCCCT	TGGTAGTTGTGGGTAAT	CCTCAAATCTACATTAAT	CCTATTAT		
common_Iguana	TTATGCAACTTTTGCCCT	TGGTAGTTGTGGGTAAT	CCTCAAATCTACATTAAT	CCTATTAT		
Lamprey_pp	CTACGCGCTCTTCGCCAT	TGATCGTGGTGACCAAG	CCCCGACCTGTACATCG	ACCCGGTCAT		
R._temporaria_p	GTACGCCGCCTTTGCCAT	GACGGTTGTGGCAAAT	CCCCGAATGCACATTG	ACCCATTAT		
R._tigrina_pp	GTACGCCGCCTTTGCCAT	GACGGTTGTGGCAAAT	CCCCGAATGCACATTG	ACCCATTAT		
X.tropicalis_pp	GTACGCCGCCTTTGCCAT	GACGGTTGTGGCAAAT	CCCCGAATGCACATTG	ACCCATTAT		
X._laevis_pp	GTACGCCGCCTTTGCCAT	GACGGTTGTGGCAAAT	CCCCGAATGCACATTG	ACCCATTAT		
Rainbow_trout_p	CTACGCGCCTTCGCCCT	CTCTGTGATCCTTGAC	CCAGCCTCCACATCAAC	CCCTCTCAT		
Cat_Fish_pp	ATATGCAGCCTTTGCCCT	GACTGTGATCATTGATT	CAACATCTATATCAACC	CAGTGAT		

	1030	1040	1050	1060	1070	1080
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	CAAGGTTTACCCATATAC	ATGGCCAAGAGCAGCACT	GTATACAACCCCTTCAT	CTACAT		
Cherry_salmon_p	GGGCACTGTGCCAGTCT	TACCTGGCCAAGAGCAG	CACCGTGTACAACCCCT	TCATCTACCT		
Green_anole_pp	AGCGACCATCCCAATGT	ACATGGCAAAATCCAG	CACATCTATAATCCAAT	CATCTACAT		
common_Iguana	AGCCACTGTCCCAATGT	ACATGGCAAAATCCAG	CACATCTATAATCCAAT	CATCTACAT		
Lamprey_pp	CGCCCACTGCCCATGT	ACCTGACCAAGAGCAG	CACGGTCTACAACCCCT	ATCATCTACAT		
R._temporaria_p	AGCCACCGTACCCATGT	ACCTGACCAAAACCAG	CACCGTCTATAATCCAAT	TATCTACAT		
R._tigrina_pp	AGCCACCGTACCCATGT	ACCTGACCAAAACCAG	CACCGTCTATAATCCAAT	TATCTACAT		
X.tropicalis_pp	AGCCACCGTACCCATGT	ACCTGACCAAAACCAG	CACCGTCTATAATCCAAT	TATCTACAT		
X._laevis_pp	AGCCACCGTACCCATGT	ACCTGACCAAAACCAG	CACCGTCTATAATCCAAT	TATCTACAT		
Rainbow_trout_p	TGCCACGCTTCCCATGT	ACCTGGCCAAGAGTAG	CACTGTCTACAACCCCT	ATCATATACGT		
Cat_Fish_pp	TGGCACAAATACCCGCA	TACCTGGCCAAGAGCAG	CACTGTCTTAAACCCCT	ATCATTTATAT		

	1090	1100	1110	1120	1130	1140
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	ATACATGAACAAACAG	-----			GTGAGCAGAAC	--
Cherry_salmon_p	CTACATGAACAAACAGT	TCCGGAGATATGCAGT	GCCCTTCCTCCTGT	GTGGGAGGGACCC		
Green_anole_pp	TTTTATGAACAAGCAGT	TCGCGATTGCCTTGTA	AGGTGTCTCCTGT	GTGGAAGAAATCC		
common_Iguana	TTTTATGAACAAGCAGT	TCGCGATTGCCTTGTA	AGGTGTCTCCTGT	GTGGAAGAAATCC		
Lamprey_pp	CTTCATGAACCGCCAGT	TCCGGGACTGCGCCGT	GCCCTTCCTGCTCT	GCGGACGCAACCC		
R._temporaria_p	TTTCATGAACAAGCAGT	TCCAAGAATGTGTCAT	TCCCTTCTGTTCT	GCGGAGGAATCC		
R._tigrina_pp	TTTCATGAACAAGCAGT	TCCAAGAATGTGTCAT	TCCCTTCTGTTCT	GCGGAGGAATCC		
X.tropicalis_pp	TTTCATGAACAAGCAGT	TCCAAGAATGTGTCAT	TCCCTTCTGTTCT	GCGGAGGAATCC		
X._laevis_pp	TTTCATGAACAAGCAGT	TCCAAGAATGTGTCAT	TCCCTTCTGTTCT	GCGGAGGAATCC		
Rainbow_trout_p	GTTTCATGAACAGACAGT	TCCGGGACTGTGCTGT	TCTTCTTGCTGT	GTGGTCTGAACCC		
Cat_Fish_pp	TTTCATGAACAGACAGT	TTCAGGGACTATGCTT	TGCCTTGCTTTTAT	GTGGAAAAACCC		

	1150	1160	1170	1180	1190	1200
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----AA-----					
Cherry_salmon_p	ATGGC-----	CATCTGAGGAG-----	GAGGGGTCAGAGATGCAGACCACTGTGGC			
Green_anole_pp	CTGTG-----	CTTCTGAGCAA-----	ACAGATGAAGA-----		TGATCTAGA	
common_Iguana	CTGTG-----	CTTCTGACCAA-----	GCTGAGGAAGA-----		TGATCTAGA	
Lamprey_pp	CTGGGCAGAGCCTTCCTCGGAG-----	TCGG-----		CGACGGC		
R._temporaria_p	TTGGG-----	CTGCTGAAAAG-----	TCAAGTT-----		CTATGGA	
R._tigrina_pp	TTGGG-----	CTGCTGAAAAG-----	TCAAGTT-----		CTATGGA	
X.tropicalis_pp	TTGGG-----	CTGCTGAAAAG-----	TCAAGTT-----		CTATGGA	
X._laevis_pp	TTGGG-----	CTGCTGAAAAG-----	TCAAGTT-----		CTATGGA	
Rainbow_trout_p	TTGGG-----	CCTCAGAACCA-----	GTGGGCTCTGAAGCTGATACCGCCCTGTC			
Cat_Fish_pp	TTGGG-----	CAGCCAAAGAAGGACGAGACTCAGACACCAATACATTAATACTACTACAGT				
	1210	1220	1230	1240	1250	1260
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----					
Cherry_salmon_p	AACCATCAACAACAAAGTCTCCCCAGTTGATCCATACAGGCATTTAGATAAATCCACTG					
Green_anole_pp	AGTC-----					TCTACCA
common_Iguana	AGTT-----					TCTACCA
Lamprey_pp	GGCC-----					TCCACGT
R._temporaria_p	AACT-----					TCTATCA
R._tigrina_pp	AACT-----					TCTATCA
X.tropicalis_pp	AACT-----					TCTATCA
X._laevis_pp	AACT-----					TCTATCA
Rainbow_trout_p	CTCCGTCAGCAAGAA-----					CCCACGA
Cat_Fish_pp	CAGC-----					AAGAACACCTCA
	1270	1280	1290	1300	1310	1320
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----					
Cherry_salmon_p	ATGGCACCAGCAGCACCAATAAAGTGTCACCTCATATTCTATGCCTTTTTTAATACATTCA					
Green_anole_pp	TTGCTCCTGCT-----					
common_Iguana	TTGCTCCTGCT-----					
Lamprey_pp	CGGCCACCAGCGTGACCC-----					
R._temporaria_p	GCGTAACGAGTGGAAC-----					
R._tigrina_pp	GCGTAACGAGTGGAAC-----					

X.tropicalis_pp	GCGTAACGAGTGGAAC-----
X._laevis_pp	GCGTAACGAGTGGAAC-----
Rainbow_trout_p	GTCTCACCCCAATGAC-----
Cat_Fish_pp	GTGTCACCATTATAACCA-----

	1330	1340	1350	1360	1370	1380
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----					
Cherry_salmon_p	TTAAATATACCTCTTAGTTTGTAAGAATATGACTTATTAATACCTTAATAAGGAGATTG					
Green_anole_pp	-----					
common_Iguana	-----					
Lamprey_pp	-----					
R._temporaria_p	-----					
R._tigrina_pp	-----					
X.tropicalis_pp	-----					
X._laevis_pp	-----					
Rainbow_trout_p	-----					
Cat_Fish_pp	-----					

	1390	1400	1410	1420	1430	1440
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----					
Cherry_salmon_p	TATGACTTTCTTACCTTGTCACAACCAAAATAACTTTAGAGTATGGTTTAAATACTACAT					
Green_anole_pp	-----CCTTCC					
common_Iguana	-----CCTTCC					
Lamprey_pp	-----TGGCGT					
R._temporaria_p	-----GCCT					
R._tigrina_pp	-----GCCT					
X.tropicalis_pp	-----GCCT					
X._laevis_pp	-----GCCT					
Rainbow_trout_p	-----AGCAACTTTCTATTTTCTTT					
Cat_Fish_pp	-----TAAGTATCACCG					

	1450	1460	1470	1480	1490	1500
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----GCTGA-----					

Cherry_salmon_p	CCAGAATGGGATGAGAGACTGGCTGATCACAGAGTATGTTACCTACAGTATGTTACAGAA
Green_anole_pp	TCAAGACGTGGCAAA-----
common_Iguana	TCAAAACGTGGCAAA-----
Lamprey_pp	CCGCGCCGGGG-CAG-----
R._temporaria_p	ACGAAACGTGGCCAA-----
R._tigrina_pp	ACGAAACGTGGCCAA-----
X.tropicalis_pp	ACGAAACGTGGCCAA-----
X._laevis_pp	ACGAAACGTGGCCAA-----
Rainbow_trout_p	GTACAATGTA---AA-----
Cat_Fish_pp	TTCATACCTAAAAACGGGTGTTTCAAATATTCAGCAATTCTCTTTCTATGTGTATTAAT

	1510	1520	1530	1540	1550	1560
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----	-----	-----	-----	-----	-----
Cherry_salmon_p	GGTATATTAACTGTGGAATCTATCGTTACTTTATGGTAAATATAAAACACTATGAACATGC					
Green_anole_pp	-----	-----	-----GTTGCTCCTGTATAG-----			
common_Iguana	-----	-----	-----GTTACTCCTATGTAGAAAAACAATTCAGCATTGCTTT			
Lamprey_pp	-----	-----	-----GTCTCCCCGAGCTAGAGCGATCGATGAACT-----			
R._temporaria_p	-----	-----	-----GTGGCTCCGGCGTAA-----			
R._tigrina_pp	-----	-----	-----GTGGCTCCGGCGTAA-----			
X.tropicalis_pp	-----	-----	-----GTGGCTCCGGCGTAA-----			
X._laevis_pp	-----	-----	-----GTGGCTCCGGCGTAA-----			
Rainbow_trout_p	-----	-----	-----GTTGCTGTTATAAAA-----			
Cat_Fish_pp	ATTGTGAATATTTACAGTGTAATTTAGCTCTGGTATTTATTGTGCTTGGATATTTATTT					

	1570	1580	1590	1600	1610	1620
	=====+	=====+	=====+	=====+	=====+	=====+
Zebrafish_pp_b	-----	-----	-----	-----	-----	-----
Cherry_salmon_p	TGTTCAAT-----ACATAATGCATGGCTTATCTATAAACAAAGCACAAATAATGAACATA					
Green_anole_pp	-----	-----	-----	-----	-----	-----
common_Iguana	GGATTGATGTACATCACAATGCTATGCACACTTTGTTCCCTGCCTGGAATCCCTGACACA					
Lamprey_pp	-----	-----	-----	-----	-----	-----
R._temporaria_p	-----	-----	-----	-----	-----	-----
R._tigrina_pp	-----	-----	-----	-----	-----	-----
X.tropicalis_pp	-----	-----	-----	-----	-----	-----
X._laevis_pp	-----	-----	-----	-----	-----	-----
Rainbow_trout_p	-----	-----	-----	-----	-----	-----
Cat_Fish_pp	GGGTGAATGGTATTGTTTAAAGACATAGTCTTTAAGGTATTGTCTTGGATTTCATAGAATA					

	1630	1640	1650	1660	1670	1680
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	-----					
Cherry_salmon_p	CAGTGT	-----			AGAATTTC	-----GTGTA
Green_anole_pp	-----					
common_Iguana	CTTACCCAGAAACAAGTCCCACTCTGTTTCAGTGAGGCTTAAATGTGTTTAGCATTCTCTGA					
Lamprey_pp	-----					
R._temporaria_p	-----C	-----			GTGCTTTTA	-----TCTCC
R._tigrina_pp	-----C	-----			GTGCTTTTA	-----TCTCC
X.tropicalis_pp	-----C	-----			GTGCTTTTA	-----TCTCC
X._laevis_pp	-----C	-----			GTGCTTTTA	-----TCTCC
Rainbow_trout_p	-----					-----TAT
Cat_Fish_pp	CAAC	-----			ATATTTTTT	-----TTTAA
	1690	1700	1710	1720	1730	1740
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	-----					
Cherry_salmon_p	AGTTTATTCAGTAATAATCATTAGAAAATAAAATATCATTATTGATTAAATGAATTCAA					
Green_anole_pp	-----					
common_Iguana	ATTTTAATCTGTCCCAACTATTGTACAATGCCCTATCATTGTCTTTGAACCTGTGTTAAT					
Lamprey_pp	-----					
R._temporaria_p	AGTATAACCGGC	-----				
R._tigrina_pp	AGTATAACCGGC	-----				
X.tropicalis_pp	AGTATAACCGGC	-----				
X._laevis_pp	AGTATAACCGGC	-----				
Rainbow_trout_p	AGTTTGATGTGCAATATTTGGTTGTAAGTGTTCAGAAATACTAATAAAAGTATAGTCCA					
Cat_Fish_pp	AGCTTATCCCTGGCTACAATTTTCTCAAGAAATTGACAAATAAAGCATTGGTTATGCCCC					
	1750	1760	1770	1780	1790	1800
	=====+=====+=====+=====+=====+=====+					
Zebrafish_pp_b	-----					
Cherry_salmon_p	GGCAAAAAAAAAAAAAAAAAAAAAAAAA-----					
Green_anole_pp	-----					
common_Iguana	GTTAACTGTCCTTTTACGGAACATTGTGTTATGGTGTATATAACATTCTGGAAATAGTTA					
Lamprey_pp	-----					



Figure 8.4 Multiple Sequence Alignment (MUSCLE)

(http://www.phylogeny.fr/version2.cgi/one_task.cgi?task_type=muscle)

Above Figure shows Multiple Sequence Alignment results of all known Parapinopsin DNA sequences mentioned in Figure 8.3. Above Figure shows 44% conserved DNA positions highlighted in black among all known Parapinopsin and newly found *X.laevis*, *R. temporaria* and *R. tigrina*.

8.3 Parapinopsin DNA Phylogenetic Tree

All Parapinopsin DNA sequences aligned in Figure 8.5 were used to build a Phylogenetic tree which determines the evolutionary relations among different Parapinopsin species.

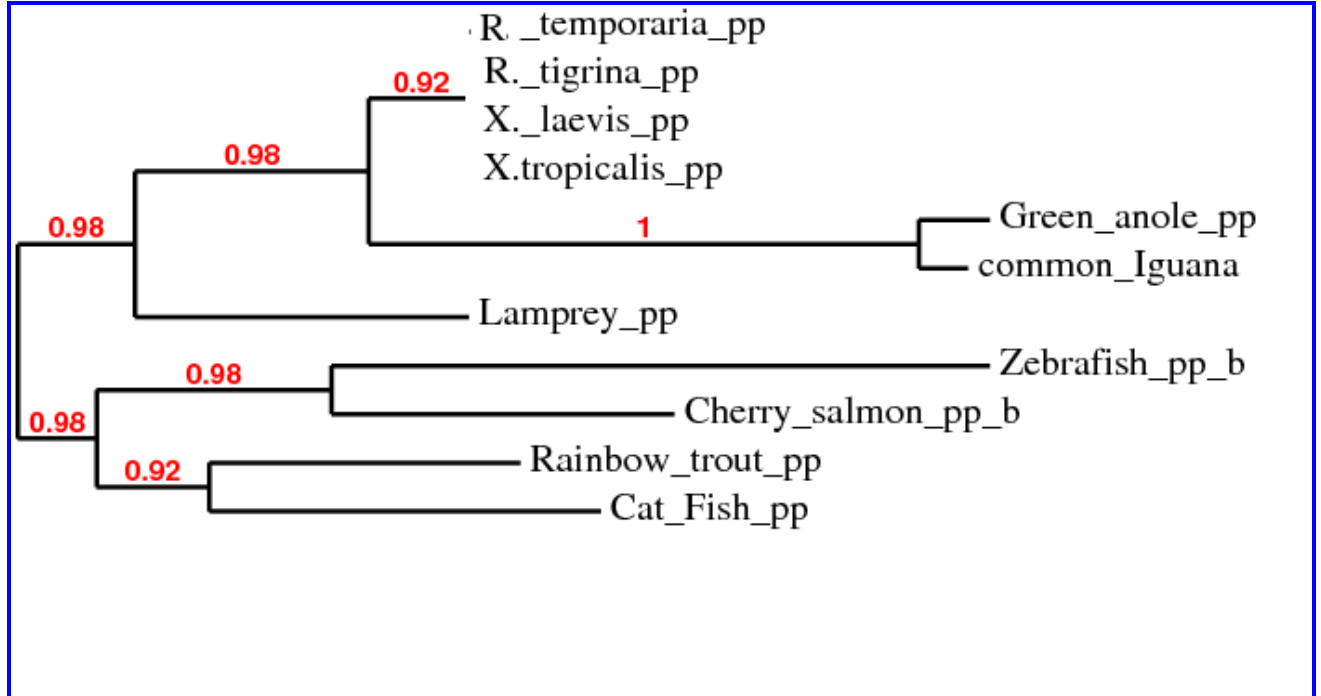


Figure 8.5 Parapinopsin DNA phylogenetic tree

(http://www.phylogeny.fr/version2_cgi/simple_phylogeny.cgi)

Figure above shows the Parapinopsin phylogenetic tree constructed by “TreeDyn 198.3” program on the basis of conserved DNA sequences. Each branch node has bootstrap values (higher bootstrap value indicates strong support for a clade). The name of each Parapinopsin is given at the right side of each cluster. All the sequences were taken from NCBI Genbank databases.

All four Amphibian species made a separate, single Parapinopsin clade on the basis of conserved DNA sequences.

8.4 Parapinopsin Study by Basic Local Alignment Similarity Tool (BLAST)

In this section Parapinopsin *X. tropicalus* Parapinopsin (NM_213665.1) was pairwise aligned with Basic Local Alignment Search Tool called “BLAST” with all known sequences available at NCBI data base e.g. GenBank, CDS translations, PDB (Protein Databank), SwissProt, PIR (Protein Information Resource) and PRF (Protein Research Foundation).

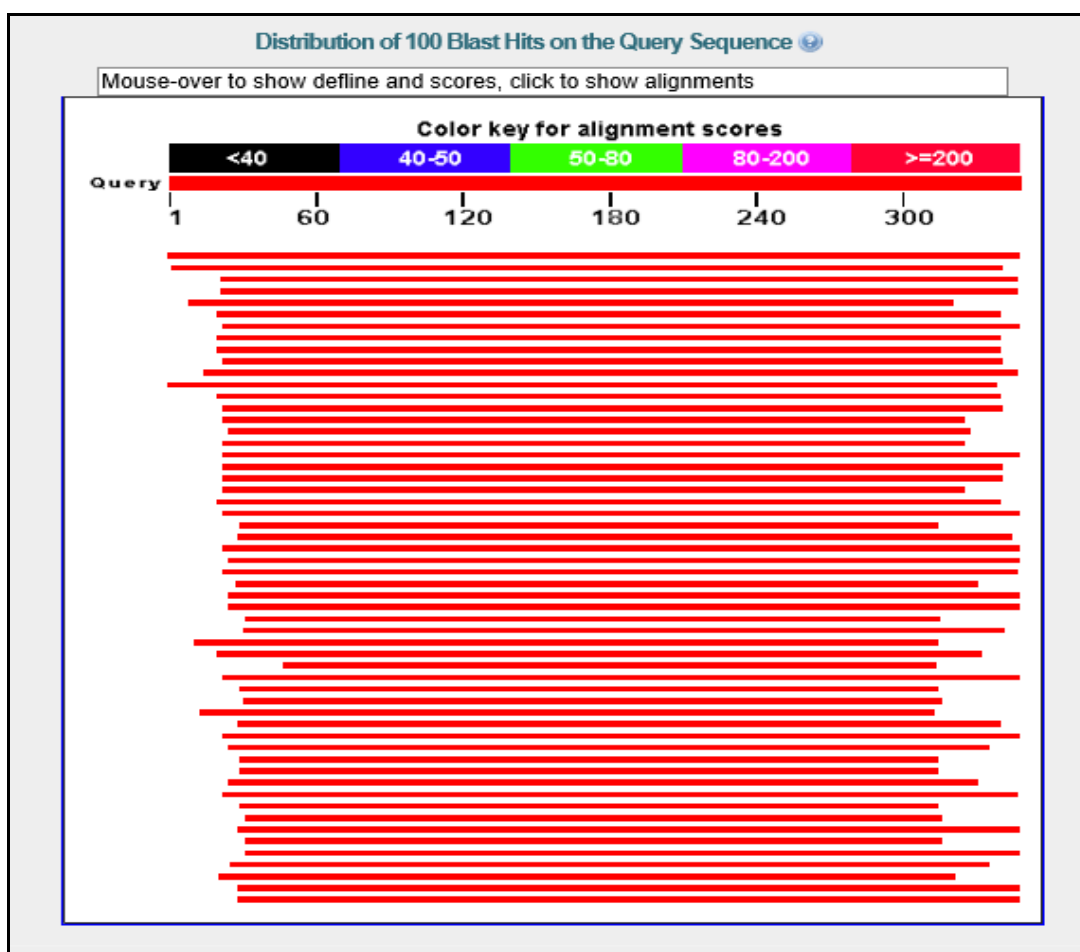


Figure 8.6 (a) Graphical representation of *X. tropicalus* Parapinopsin (NM_213665.1) BLAST results

Figure above shows the *X. tropicalus* Parapinopsin 348 amino acid residues (query sequence) matches with other known sequences available in Databank shown in red bars. Alignment scores are represented by colour key. Maximum alignment score is >200 is shown in red and minimum alignment score is <40 shown in black. All the rest of alignment scores are in between 40-200 represented in blue, green and pink colour code respectively.

The red bars in above Figure is explained in Figure 8.6 (b) representing the names of different species whose sequences match with *X. tropicalus* Parapinopsin.

Description	Max score	Total score	Query cover	E value	Ident	Accession
parapinopsin [Xenopus (Silurana) tropicalis]	342	342	100%	2e-112	100%	NP_998830.1
PREDICTED: parapinopsin-like [Lepisosteus oculatus]	256	256	97%	6e-79	67%	XP_006631190.1
parapinopsin [Iguana iguana]	238	238	93%	7e-72	69%	BAM28745.1
parapinopsin [Anolis carolinensis]	238	238	93%	1e-71	68%	NP_001280060.1
bistable UV pigment [Leithenteron camtschaticum]	229	229	89%	1e-68	74%	BAD13381.1
PREDICTED: parapinopsin-like [Maylandia zebra]	227	227	91%	8e-68	56%	XP_004548917.1
PREDICTED: pinopsin-like [Latimeria chalumnae]	226	226	93%	3e-67	45%	XP_005987605.1
PREDICTED: parapinopsin-like [Haplochromis burtoni]	226	226	91%	4e-67	56%	XP_005948736.1
PREDICTED: parapinopsin-like [Pundamilia nyererei]	224	224	91%	9e-67	56%	XP_005741571.1
PREDICTED: parapinopsin-like [Oreochromis niloticus]	225	225	91%	9e-67	59%	XP_003448342.1
parapinopsin [Ictalurus punctatus]	224	224	95%	2e-66	61%	NP_001187002.1
parapinopsin [Oncorhynchus mykiss]	223	223	97%	3e-66	58%	NP_001117697.1
PREDICTED: parapinopsin-like [Oreochromis niloticus]	222	222	91%	8e-66	56%	XP_003459791.1
PREDICTED: parapinopsin-like [Haplochromis burtoni]	223	223	91%	8e-66	58%	XP_005950795.1
PREDICTED: pinopsin-like [Chelonia mydas]	223	223	87%	8e-66	49%	XP_007057799.1
PREDICTED: pinopsin-like [Lepisosteus oculatus]	223	223	87%	1e-65	49%	XP_006641022.1
PREDICTED: pinopsin-like [Chrysemys picta bellii]	222	222	87%	2e-65	49%	XP_005298206.1
P opsin [Anolis carolinensis]	221	221	93%	2e-65	46%	AAD32622.1
PREDICTED: parapinopsin-like [Pundamilia nyererei]	221	221	91%	3e-65	58%	XP_005732739.1
PREDICTED: parapinopsin-like [Maylandia zebra]	221	221	91%	4e-65	58%	XP_004558979.1
PREDICTED: pinopsin-like [Pelodiscus sinensis]	221	221	87%	5e-65	48%	XP_006138790.1
PREDICTED: parapinopsin-like [Neolamprolochus brichardi]	220	220	91%	5e-65	56%	XP_006783555.1
pinopsin [Iguana iguana]	220	220	93%	1e-64	47%	BAM28747.1
GJ23988 [Drosophila virilis]	221	221	81%	1e-63	22%	XP_002053610.1
PREDICTED: pinopsin-like [Ciona intestinalis]	217	217	90%	1e-63	41%	XP_002119963.1
pinopsin [Uta stansburiana]	217	217	93%	2e-63	47%	AAZ79905.1
PREDICTED: pinopsin-like [Columba livia]	216	216	92%	2e-63	45%	XP_005512018.1
PREDICTED: opsin-VA-like [Anolis carolinensis]	216	216	93%	4e-63	44%	XP_003221090.1
PREDICTED: parapinopsin-like [Poecilia formosa]	215	215	87%	4e-63	58%	XP_007557772.1
deep brain photoreceptor protein pinopsin [Bufo japonicus]	215	215	92%	5e-63	46%	AAF12820.1
RecName: Full=Pinopsin; AltName: Full=Pineal gland-specific opsin; Short=P-opsin; Short=Pineal opsin [Columba livia]	215	215	92%	5e-63	45%	P51476.1
PREDICTED: beta-1 adrenergic receptor [Myotis brandtii]	215	215	81%	6e-63	21%	XP_005874624.1
PREDICTED: somatostatin receptor type 2 [Sorex araneus]	216	216	89%	7e-63	23%	XP_004618085.1
unnamed protein product [Oncorhynchus mykiss]	219	219	87%	7e-63	21%	CDQ72256.1
parapinopsin B [Oncorhynchus masou]	215	215	89%	7e-63	58%	BAN14822.1
PREDICTED: alpha-1A adrenergic receptor isoform X1 [Ficedula albicollis]	218	218	76%	8e-63	19%	XP_005058090.1
PREDICTED: parapinopsin-like [Astyanax mexicanus]	214	214	93%	1e-62	64%	XP_007237135.1
GF17441 [Drosophila ananassae]	219	219	81%	1e-62	22%	XP_001952989.1
dopamine D1/beta receptor [Branchiostoma lanceolatum]	215	215	82%	1e-62	21%	CAA06536.1
PREDICTED: somatostatin receptor type 5-like [Astyanax mexicanus]	217	217	86%	2e-62	23%	XP_007240683.1
rCG41589 [Rattus norvegicus]	217	217	89%	2e-62	20%	EDM01084.1

PREDICTED: parapinopsin-like [Maylandia zebra]	221	221	91%	4e-65	58%	XP_004558979.1
PREDICTED: pinopsin-like [Pelodiscus sinensis]	221	221	87%	5e-65	48%	XP_006138790.1
PREDICTED: parapinopsin-like [Neolamprolochs brichardi]	220	220	91%	5e-65	56%	XP_006783555.1
pinopsin [Iguana iguana]	220	220	93%	1e-64	47%	BAM28747.1
GJ23988 [Drosophila virilis]	221	221	81%	1e-63	22%	XP_002053610.1
PREDICTED: pinopsin-like [Ciona intestinalis]	217	217	90%	1e-63	41%	XP_002119963.1
pinopsin [Uta stansburiana]	217	217	93%	2e-63	47%	AAZ79905.1
PREDICTED: pinopsin-like [Columba livia]	216	216	92%	2e-63	45%	XP_005512018.1
PREDICTED: opsin-VA-like [Anolis carolinensis]	216	216	93%	4e-63	44%	XP_003221090.1
PREDICTED: parapinopsin-like [Poecilia formosa]	215	215	87%	4e-63	58%	XP_007557772.1
deep brain photoreceptor protein pinopsin [Bufo japonicus]	215	215	92%	5e-63	46%	AAF12820.1
RecName: Full=Pinopsin; AltName: Full=Pineal gland-specific opsin; Short=P-opsin; Short=Pineal opsin [Columba livia]	215	215	92%	5e-63	45%	P51476.1
PREDICTED: beta-1 adrenergic receptor [Myotis brandtii]	215	215	81%	6e-63	21%	XP_005874624.1
PREDICTED: somatostatin receptor type 2 [Sorex araneus]	216	216	89%	7e-63	23%	XP_004618085.1
unnamed protein product [Oncorhynchus mykiss]	219	219	87%	7e-63	21%	CDQ72256.1
parapinopsin B [Oncorhynchus masou]	215	215	89%	7e-63	58%	BAN14822.1
PREDICTED: alpha-1A adrenergic receptor isoform X1 [Ficedula albicollis]	218	218	76%	8e-63	19%	XP_005058090.1
PREDICTED: parapinopsin-like [Astyanax mexicanus]	214	214	93%	1e-62	64%	XP_007237135.1
GF17441 [Drosophila ananassae]	219	219	81%	1e-62	22%	XP_001952989.1
dopamine D1/beta receptor [Branchiostoma lanceolatum]	215	215	82%	1e-62	21%	CAA06536.1
PREDICTED: somatostatin receptor type 5-like [Astyanax mexicanus]	217	217	86%	2e-62	23%	XP_007240683.1
rCG41589 [Rattus norvegicus]	217	217	89%	2e-62	20%	EDM01084.1
pinopsin [Podarcis siculus]	213	213	93%	4e-62	44%	AAY34940.2
PREDICTED: pinopsin-like [Zonotrichia albicollis]	213	213	89%	4e-62	45%	XP_005493623.1
PREDICTED: tachykinin-like peptides receptor 86C-like isoform X1 [Ceratitis capitata]	217	217	81%	5e-62	23%	XP_004529189.1

Figure 8.6 (b) *X.tropicalis* BLAST results

Figure above represent the alignment results of different protein sequences with *X.tropicalis*

Column 1 represents the name of protein sequences

Column 2, 3 and 4 represents the total/maximum score on the basis of sequence similarities and Query cover respectively

Column 5 represents the expect-value (e-value); the number of time the particular sequence hit the database by chance. With the increase of matching score the E-value decreases

Column 6 represents the % identity of query sequence with database sequence

Column 7 represents the accession numbers

8.5 BLAST Results Phylogenetic Analysis by Neighbour Joining Method

X.tropicalis Parapinopsin BLAST results mentioned in section 8.4 were used to build phylogenetic tree by neighbour joining method.

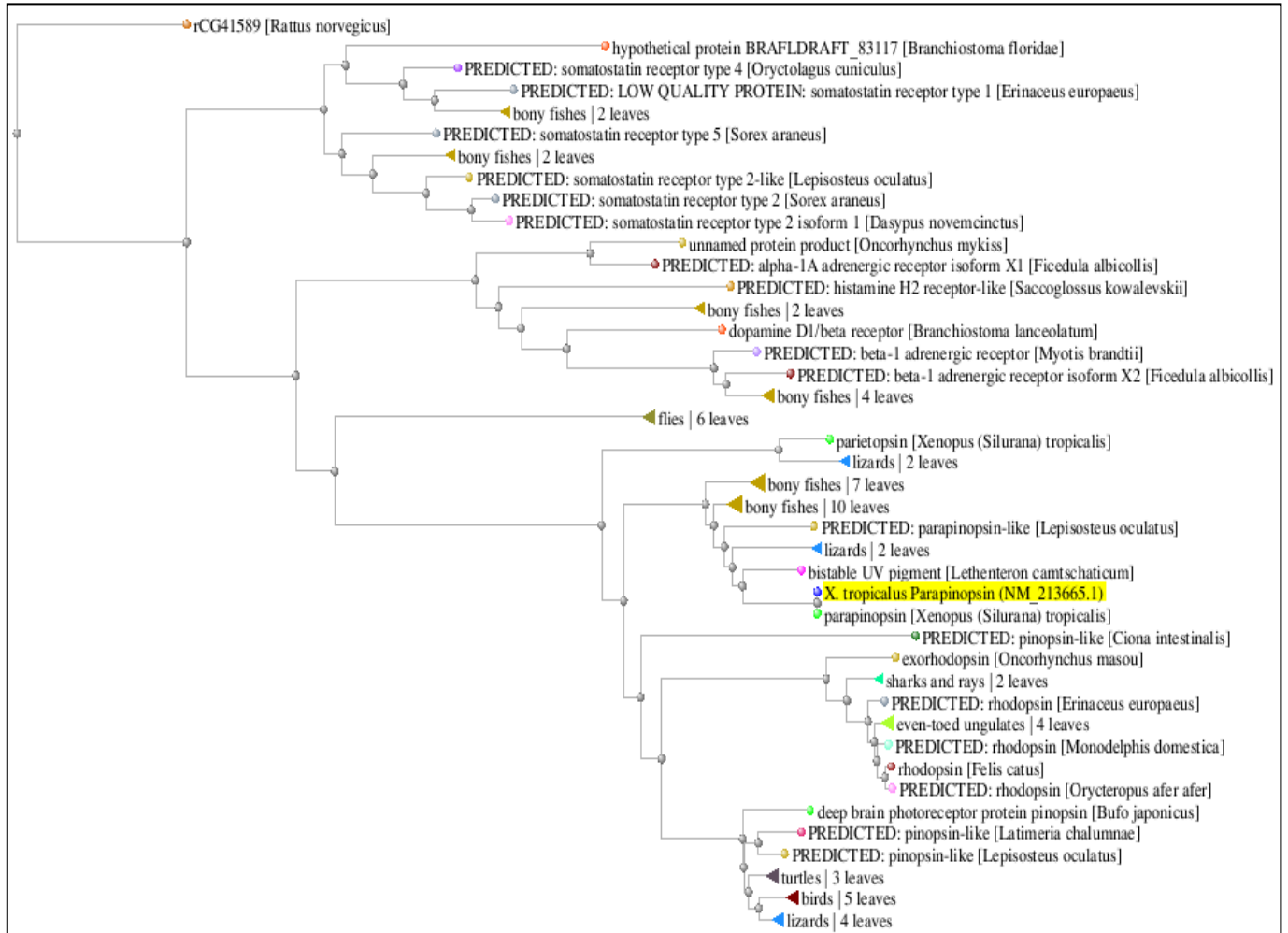


Figure 8.7 Opsin based phylogenetic tree constructed by neighbour joining method

Phylogenetic tree in above Figure based on BLAST results, shows position of *X.tropicalis* highlighted in yellow among different visual, non-visual opsin and other Parapinopsin. The name of each opsin/Parapinopsin is given at the right side of each cluster.

The *X.tropicalis* Parapinopsin is positioned in the middle of lower vertebrate visual and higher vertebrate non-visual and visual opsins.

8.6 Discussion

Bioinformatics is Computational Molecular Biology. The aim of bioinformatics is the delivery of fast and efficient molecular data and availability of user friendly tools to determine the DNA to protein interaction and in establishing a relationship among diverse organisms through phylogenetic analysis. The bioinformatics tools are like a fishing fleet acting in a large ocean (the database) and are used to mine information which is hidden to us. The total number of nucleotide sequences of different species at GenBank (<http://www.ncbi.nlm.nih.gov>) database is “260000”, submitted by individual labs or from groups through whole genome shotgun sequencing (WGS). GenBank covers the worldwide nucleotide sequences by exchanging data on a daily basis from the European Nucleotide Archive (ENA) and the DNA Data Bank of Japan (DDBJ). GenBank provides links to major DNA, protein sequences, taxonomy, genome, mapping, protein structure and PubMed. GenBank similarity searches are found out by “BLAST” (D. A. Benson *et al.*, 2013). The database size doubles after every 12 months (Luscombe *et al.*, 2001).

All Parapinopsin nucleotide and protein sequences, along with visual and non-visual opsins, were taken from Genbank (<http://www.ncbi.nlm.nih.gov>) and used for multiple sequence alignment (MSA) available at EBI (European Biotechnology Institute) called clustalW MSA (Larkin *et al.*, 2007; Goujon *et al.*, 2010 and McWilliam *et al.*, 2013). The conserved amino acid sequences shared by different Parapinopsins are between ~40-60% (Table 8.2) and also reported in literature by Tsukamoto *et al.*, 2009, Shichida and Matsuyama, 2009).

All Parapinopsin proteins sequences were aligned through Multiple Sequence Alignment CLUSTALW (Thompson *et al.*, 1994), the most widely used program for sequence alignment. Alignment results showed the maximum sequence similarity between *X. tropicalis* Parapinopsin and Lamprey Parapinopsin which is 66.76% and between *X. tropicalis* Parapinopsin and Human Rhodopsin protein is 34.48%. Hence Parapinopsin

sequence conservation is gradually decreased from lower vertebrate non-visual opsins to modern visual opsins.

Collection of data and alignment of sequences are important steps to build a phylogenetic tree. Hence all Parapinopsin nucleotide sequences were taken from Genbank available at NCBI (given in Figure 8.3) and aligned. The modern, fast and highly efficient computer based alignment program “MUSCLE” (Multiple sequence comparison by log-expectation) is used for nucleotide sequence alignment (Edgar, 2004). The common Parapinopsin nucleotide sequences were highlighted in black (Figure 8.4).

The “Phylogeny.fr” (Dereeper *et al.*, 2008) is a free accessed, easy to use web service for Phylogenetic tree construction and establishes relationships among different molecular sequences (http://www.phylogeny.fr/version2_cgi/simple_phylogeny.cgi).

The approach used to construct a phylogenetic tree during the current study was “Maximum Likelihood”, called ML (Felsenstein, 1981; Guindon and Gascuel, 2003). The algorithm used for ML simply creates a fast “distance based tree” and then improve the likelihood of tree along with branch length. The phylogenetic trees (Figures 8.2 and 8.4) were produced by using fastest PHYLIM package of the ML program.

The Parapinopsin nucleotide phylogenetic tree represents the evolution of modern opsin from Parapinopsin where human Rhodopsin was used an out-group allows to root the tree. Maximum Likelihood was used to determine the substitutions among different sequences to establish ancestral relationships.

To best of my knowledge the *X. laevis*, *R. temporaria* and *R.tigrina* Parapinopsin sequences have not been reported and were found for the first time during the current project, were used to construct a phylogenetic tree along with other known Parapinopsins. All four frog species along with green anole and common iguana share a common clade (Figure 8.4).

BLAST search was carried out to find local alignment for similarity searches and to find neighbours of Parapinopsin through a phylogenetic tree (Boratyn *et al.*, 2012; Altschul *et al.*, 1997 and Schäffer *et al.*, 2001). A phylogenetic tree was constructed through “Neighbour Joining” method. *X. tropicalis* Parapinopsin query sequence is highlighted yellow (Figure 8.6) and shows a position in between modern vertebrate visual/non-visual opsins and invertebrate visual opsin.

Chapter 9 Discussion

9.1 Discussion

Parapinopsin is a G-protein coupled receptor (GPCR) and a UV sensitive non-visual protein expressed in pineal organs of lower vertebrates (Koyanagi *et al.*, 2004; Terakita 2005; Wada *et al.*, 2012; Koyanagi and Terakita, 2013). Parapinopsin was studied for the first time by Seth Blackshaw and Solomon H. Snyder, 1997 and shows ~60% similarity with bovine rhodopsin (Tsukamoto *et al.*, 2009). Parapinopsin is a key pigment in understanding the evolution of vertebrate visual pigments. The vertebrate visual pigment has a counterion, aspartic acid, at position E113 while the invertebrate visual pigment has a counterion, aspartic acid, at position E118. Parapinopsin has a counterion at both E113 and E118 positions. The ability to activate G protein is much less in Parapinopsin, like invertebrate visual pigments. These facts suggest that the vertebrate visual pigment evolved from an ancestral vertebrate non-visual pigment like Parapinopsin (Yamashita *et al.*, 2011; Tsukamoto and Terakita, 2010).

The aim of the project was to study Parapinopsin in amphibians (*Xenopus tropicalis*, *Xenopus laevis*, *Rana temporaria* and *Rana tigrina*) because little is known about this protein and in amphibians only *X. tropicalis* Parapinopsin cDNA sequence has been reported at NCBI. The project was divided into 5 parts: 1) the study of the Parapinopsin coding sequence, 2) SSP-PCR (strand specific primers- Polymerase chain reaction) based analysis of amphibian's cDNA and genomic DNA, 3) Peptide synthesis and Western Blot analysis of Parapinopsin protein, 4) Northern Blot analysis for Parapinopsin expression in *X. laevis* melanophores, 5) a Bioinformatics based study of Parapinopsin and 6) Expression studies of Parapinopsin protein in the GPCR specific expression vectors, pASK75-GPCR, pASK75-GPCR-GFP and pBAD33.

The results of the first part of the project revealed novel findings. The Parapinopsin cDNA study of *X. laevis* tadpole showed 100% homology with *X. tropicalis* Parapinopsin cDNA. These results were according to expectation, and reported literature by Hirsch *et al.*, 2002; Burki 1985; RO & Hillis 1990. Due to being close relatives, most of the

sequences are conserved in *X. tropicalis* and *X. laevis*. The adult frog *R. tigrina* and tadpole *R. temporaria* Parapinopsin cDNA sequences showed ~99.9% similarity with *X. tropicalis* Parapinopsin cDNA. The 0.1% dissimilarity is due to a point mutation from G to A at the 2nd position in the 156th codon. Hence *R. temporaria* and *X. tropicalis* Parapinopsin gene are orthologous (homologous genes present in different species) genes. Orthologous genes show some variation during speciation event divergence (Jenson, 2001). This point mutation is very important as it creates a stop codon UGG (tryptophan/W)→ UAG and hence protein truncation. Due to the importance of a point mutation, the results were further confirmed by SSP-PCR analysis.

Genomic DNA (gDNA) sequence study results revealed a heterozygosity in *R. temporaria* at the position 2/156th codon, where a point mutation was found at the cDNA level. The sequencing results for 3 samples (1, 2 & 3) revealed heterozygosity in *R. temporaria* gDNA at 2/156th position and showed equal ratios for both normal and mutant alleles, (table 3.5). For 4 samples (4, 5, 6 and 7) sequencing results showed two peaks of unequal heights where the ratio for mutant to normal allele was high. Initially the gDNA study was conducted on sets of three tadpoles per reaction. Later on “artificial heterozygotes” and “single tadpole gDNA study” was conducted and used as a control to answer the question: why were unequal mutant to normal ratios present at the heterozygous locus? Single tadpole gDNA sequence study results of five samples were like samples 4, 5, 6, 7 and 8 (table 3.5). Artificial heterozygotes were produced by mixing cloned cDNAs of *R. temporaria* and *X. laevis* at 50:50 ratios and confirmed the sequencing results. The *R. temporaria* gDNA study showed genetic variability at 2/156th codon in different *R. temporaria* samples. Genetic variability is common in amphibians. The possible reasons for this intraspecies mutations or genetic variability may be due to a slow rate of evolution at the gene level, history of the population and a small population size which produces a high frequency of inbreeding. Climatic extremism is also very important e.g. human disturbance, and global warming results in habitat fragmentation (Chen *et al.*, 2012; Fu *et al.*, 2001).

The *R. temporaria* gDNA results may be interpreted in terms of loss of heterozygosity or loss of genetic diversity on the basis of high mutant to normal allele ratios in most of the samples. Amphibians are at a higher risk because they are an endangered species. Loss of genetic diversity may result in reduction of population size, or may be the reduction in population size is due to loss of genetic variability. One assumption may be that the higher rate of genetic homology may cause the reduction in fitness level due to dramatic change in environment (Chen *et al.*, 2012). Matsuba *et al.*, in 2008 found sex reversal, and undifferentiated sex chromosomes are the common characteristics of *Rana temporaria*. In *R. temporaria*, the increased temperature (>32°C) results in female's turning into males.

Western blotting was used to determine more about the size of the protein produced in *Rana*. Two peptide regions were selected at N and C termini. Western Blotting results support the cDNA study results with a 37 kDa protein band of Parapinopsin in the *X. laevis* protein sample which is in accord with the size of green iguana Parapinopsin (Wada *et al.*, 2012) ~38kDa (37.9kDa).

Bioinformatics software was used to study evolutionary genetics. All known Parapinopsin cDNA and amino acid sequences were used to generate phylogenetic trees. The phylogenetic tree analysis revealed the evolution of modern opsins from Parapinopsin as reported by Yamashita *et al.*, 2011.

RT-PCR (reverse transcriptase- PCR) based Parapinopsin studies were conducted by Yamashita and his colleagues in 2007 to check the expression of Parapinopsin in various tissues of lamprey e.g. brain, pineal, eye and skin. They found that Parapinopsin is only expressed in pineal tissues. During the current project a similar approach was used. Northern Blot studies were conducted on *X. laevis* melanophores to look for Parapinopsin expression. Results indicated the absence of Parapinopsin expression in *X. laevis* melanophores.

For protein expression studies the entire coding sequence of Parapinopsin gene of *X. tropicalis* from the amino acid sequence was chemically synthesized. The chemically

synthesized codon optimized gene was tailored successfully by adding 110 nucleotides at the 5' end and restriction sites at both 5' and 3' ends. The modified Parapinopsin gene was cloned into specific G-protein coupled receptors plasmids e.g. pASK75-GPCR, pASK75-GPCR-GFP and pBAD33 to over express the protein in *E. coli* by using the same technique as used by Link *et al.*, 2008; Skretas & Georgiou, 2008 and Skretas, 2012. The aim of this study was to overexpress and purify the protein for functional studies as well as structural studies using X-ray crystallography and high field NMR. The current aim is to make sure that the plasmid is correctly engineered in order to overexpress the Parapinopsin protein.

9.2 Further work

The DNA sequence encoding the N-terminal amino acids (1-39) of Parapinopsin from *X. laevis* and the two *Rana* species was not obtained because none of the PCR primers based on the *X. tropicalis* sequence for this region produced a PCR product. This could be attributable to incompatibility between PCR primer pairs or to genuine differences from the *X. tropicalis* sequence in this region. The N-terminal region of the protein is predicted to be located in the cytoplasm and might therefore be involved in interactions with other intracellular proteins participating in transmission of the light signal so N-terminal sequence differences in Parapinopsin could modify the downstream signalling pathway by affecting the binding interactions with molecules of the signalling pathway. Whole genome sequencing is required to establish the remainder of the Parapinopsin gene sequence and the intron/exon boundaries and should also reveal the promoter and other regions involved in regulation of Parapinopsin expression. *X. laevis* is allotetraploid and its Parapinopsin gene sequence, determined in the present work, is based on a cDNA clone originating from one mRNA molecule so the possibility exists of variant forms of Parapinopsin encoded by other copies of the gene within the cell.

The results revealing nonsense mutations in the *Rana* Parapinopsin genes would predict either synthesis of lower levels of Parapinopsin or no Parapinopsin at all. Comparisons of circadian rhythms and UV sensitivities between heterozygous and homozygous mutant

individual *Rana* species using animals and cell cultures might throw some light on the physiological function of Parapinopsin.

9.3 Conclusion

The Parapinopsin gene is polymorphic in *Rana temporaria*, having two copies of the gene, one with A and other with G at the 2/156th codon. The Parapinopsin gene is not polymorphic in *X.tropicalis*, *X. laevis* and *Rana tigrina*. *X. laevis* cDNA and genomic DNA studies showed wild type Parapinopsin similar to that reported for *X.tropicalis* Parapinopsin at NCBI, both copies of the genes having G. It was concluded from *Rana tigrina* and *R. temporaria* cDNA studies, that there is a point mutation at 2nd position of 156th codon that introduces A and results in a stop codon and hence protein truncation. The SSP-PCR (Strand Specific Primer-PCR) results were in accordance with the Parapinopsin sequence studies.

cDNA, gDNA and Bioinformatics studies revealed that normal Parapinopsin protein comprises seven transmembrane helices (365 amino acids) in *X.tropicalis* and *X.laevis* but the truncated protein of showed that there is no Parapinopsin protein synthesized by *R. temporaria* and *R. tigrina* while normal Parapinopsin protein (37kDa) is present in *X.laevis*.

Parapinopsin is only expressed in the pineal and pineal related organs, and photoreceptive *X.laevis* melanophore cell line study showed the absence of Parapinopsin in melanophore

The Parapinopsin Phylogenetic study results concluded that Parapinopsin in four frog species, *X.tropicalis*, *X.laevis*, *R. temporaria* and *R. tigrina*, has a common ancestor, as they make a single clade due to there being 99-100% homology. This high level of homology has not been reported earlier among already known Parapinopsins.

The frog, *Rana*, species is important in evolutionary studies. *R. temporaria* and *R.tigrina* provides an evolutionary link between invertebrates and higher vertebrate opsins because Parapinopsin is not present in higher vertebrates.

9.4 Future Development

Future developments for the project may include the overexpression and purification of Parapinopsin protein so that it can be used for structural biological studies. Crystallization conditions will be screened and if they give rise to well diffracting crystals, then X-ray diffraction data will be collected to characterize their space group and cell dimensions, followed by synchrotron X-ray data collection at the Diamond Synchrotron for structure solution. NMR structural studies will also be pursued as has been done for other GPCRs in Gerhard Wagner's laboratory at Harvard Medical School (Royal Soc. of Chemistry, NMR meeting, Cambridge, April 2014).

The study of adult *R. temporaria* (common garden frog) population genetics is also very important as the literature reviewed about the specie gives information about intraspecific genetic variation, genome size, variation between male and female frogs. Parapinopsin study of the adult frog population would be important in terms of Parapinopsin heterogeneity at the 2/156th codon at the genomic DNA level and the mutation at the cDNA level.

Generation of more specific monoclonal antibodies instead of Polyclonal antibodies and these of pineal tissue of adult frog, would provide more accurate results.

Environmental conditions, particularly temperature, are also very important in a Frog's life. All *R. temporaria* tadpoles used during current studies were taken from the natural environment. There is a need for comparative molecular based Parapinopsin study of the common frog grown under controlled and natural environment.

References

- Akashi, H., N. Osada, and T. Ohta. (2012). "Weak Selection and Protein Evolution." Genetics. 192, 15–31.
- Alanazi, M., Z. Abduljaleel, W. Khan, A. S. Warsy, M. Elrobh, Z. Khan, A. Al Amri, M. D. Bazzi. (2011) "In Silico Analysis of Single Nucleotide Polymorphism (SNPs) in Human b-Globin Gene". PLoS ONE. 6(10): e25876
- Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman. (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402
- Anisimova, M. and O. Gascuel. (2006). "Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative". Syst Biol. 55(4):539-52.
- Arendt, D. and J. Wittbrodt. (2001). "Reconstructing the eyes of Urbilateria". Philos Trans R Soc Lond B Biol Sci. 29;356(1414):1545-63.
- Arendt, D. (2003). "Evolution of eyes and photoreceptor cell types". Int J Dev Biol. 47(7-8):563-71
- Arendt, D., K. Tessmar-Raible, H. Snyder, A. W. Dorresteijn, J. Wittbrodt. (2004). "Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain". Science. 29;306 (5697): 869-71.
- Bellingham, J., D. J. Wells, R. G. Foster. (2003). "*In silico* characterisation and chromosomal localisation of human RRH (peropsin)--implications for opsin evolution". BMC Genomics. 24; 4 (1):3
- Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, E. W. Sayers. (2013). "GenBank". Nucleic Acids Res. 41(Database issue):D36-42.

- Bianco, I. H. and S. W. Wilson. (2009). "The habenular nuclei: a conserved asymmetric relay station in the vertebrate brain". Philos Trans R Soc Lond B Biol Sci 12;364 (1519):1005-20
- Bjarnadóttir, T. K., D. E. Gloriam, S. H. Hellstrand, H. Kristiansson, R. Fredriksson, H. B. Schiöth. (2006). "Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse". Genomics 88 (3): 263–73
- Blackshaw, S. and S. H. Snyder. (1997). "Parapinopsin, a novel catfish opsin localized to the parapineal organ, defines a new gene family". Neuroscience. 1; 17(21) 8083-92.
- Bonin, A., P. Taberlet, C. Miaud, F. Pompanon. (2006). "Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*)". Mol Biol Evol. 23(4):773-83
- Boratyn, G. M., A. A. Schäffer, R. Agarwala, S. F. Altschul, D. J. Lipman, T. L. Madden. (2012). "Domain enhanced lookup time accelerated BLAST", Biology Direct 7:12
- Bottema, C. D. and S. S. Sommer. (1993). "PCR amplification of specific alleles: rapid detection of known mutations and polymorphisms". Mutat Res. 288(1):93-102.
- Bottema, C. D. and S. S. Sommer. (1993). "PCR amplification of specific alleles: rapid detection of known mutations and polymorphisms". Mutat Res. 288(1):93-102.
- Briggs, R. and King, T. J. (1952). "Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs". Proc Natl Acad Sci U S A. 38(5):455-63
- Briscoe, A. D. and L. Chittka. (2001). "The evolution of color vision in insects". Annu Rev Entomol. 46:471-510.
- Brockman, W., P. Alvarez, S. Young, M. Garber, G. Giannoukos, W. L. Lee, C. Russ, E. S. Lander, C. Nusbaum, D. B. Jaffe. (2008). "Quality scores and SNP detection in sequencing-by-synthesis systems". Genome Res. 18(5):763-70

- Brown, D. D. (2004). "A tribute to the *Xenopus laevis* oocyte and egg". J Biol Chem. 29;279(44):45291-9
- Brylinski, M., L. Konieczny, P. Czerwonko, W. Jurkowski, I. Roterman. (2005). "Early-stage folding in proteins (in silico) sequence-to-structure relation". J Biomed Biotechnol. 2: 65–79
- Burki, E. (1985). "The expression of creatine kinase isozymes in *Xenopus tropicalis*, *Xenopus laevis*, and their viable hybrid". Biochemical Genetics,. 23(1-2): p. 73-88.
- Cano, J. M., M. H. Li, A. Laurila, J. Vilkkilä, J. Merilä. (2011). "First-generation linkage map for the common frog *Rana temporaria* reveals sex-linkage group". Heredity 107, 530–536
- Castresana, J. (2000). "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis". Mol Biol Evol. 17(4):540-52
- Chen, S., Y. Zhang, X. Wang, J. Sun, Y. Xue. (2012). "Extremely low genetic diversity indicating the endangered status of *Ranodon sibiricus* (Amphibia: Caudata) and implications for phylogeography". PLOS ONE, 7 (3), e33378
- Chin, E. L., C. da Silva, M. Hegde. (2013). "Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations". BMC Genet. 14: 6
- Choe, H. W., J. H. Park, Y.J. Kim, O. P. Ernst. (2011). "Transmembrane signaling by GPCRs: insight from rhodopsin and opsin structures". Neuropharmacology. 60(1):52-7
- Curtis, P. C., J. L. Thomas, N. R. Phillips, R. K. Roby. (2010) "Optimization of primer-specific filter metrics for the assessment of mitochondrial DNA sequence data". Mitochondrial DNA. 21(6):191-7
- de Sa, R. and D. Hillis. (1990). "Phylogenetic relationships of the pipid frogs *Xenopus* and *Silurana*: and Integration of Ribosomal DNA and morphology". Mol Biol Evol 7(4): 365-376.

Debreceni, K., Z. Fejér, A. Szél, P. Röhlich, T. Görös, B. Vigh. (1998). "Photoreceptors sensitive for various wave-lengths in the pineal complex and retina of reptiles immunocytochemical localization of opsins". Neurobiology (Budapest) 6, 463-465.

Delgado, M. J., A. L. Alonso-Gómez, B. Gancedo, N. de Pedro, A. Valenciano, M. Alonso-Bedate. (1993). "Serotonin N-acetyltransferase (NAT) activity and melatonin levels in the frog retina are not correlated during the seasonal cycle". Gen Comp Endocrinol. 92(2):143-50

Dereeper, A., S. Audic, J. M. Claverie, G. Blanc. (2010). "BLAST-EXPLORER helps you building datasets for phylogenetic analysis". BMC Evol Biol. 10:8.

Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie, O. Gascuel. (2008). "Phylogeny.fr: robust phylogenetic analysis for the non-specialist" Nucleic Acids Research. 36 W465-9

Dewi, F. R. P. and F. Fatchiyah. (2013). "Methylation impact analysis of erythropoietin (EPO) Gene to hypoxia inducible factor-1 α (HIF-1 α) activity". Bioinformation 9(15): 782-787

Do, H. and A. Dobrovic. (2009). "Limited copy number-high resolution melting (LCN-HRM) enables the detection and identification by sequencing of low level mutations in cancer biopsies". Mol Cancer. 8;8:82

Ebrey, T. and Y. Koutalos. (2001). "Vertebrate photoreceptors". Prog Retin Eye Res. 20(1):49-94

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput". Nucleic Acids Res. 19;32(5):1792-7

Efron, B., E. Halloran, S. Holmes. (1996). "Bootstrap confidence levels for phylogenetic trees". Proc Natl Acad Sci U S A. 12;93(23):13429-34.

Elsdale, T. R., J. B. Gurdon, M. Fischberg. (1960). "A description of the technique for nuclear transplantation in *Xenopus laevis*". J Embryol Exp Morphol. 8:437-44

Engbretson, G. A. and V. H. Hutchison. (1976). "Parietalectomy and thermal selection in the lizard *Sceloporus magister*". J. Exp. Zool. 198, 29-38.

Fejér, Z., P. Röhlich, A. Szél, C. Dávid, A. Zádori, M. J. Manzano, B. Vigh. (2001). "Comparative ultrastructure and cytochemistry of the avian pineal organ". Microsc. Res. Techn. 53, 12-24.

Felsenstein, J. (1981). "Evolutionary trees from DNA sequences: a maximum likelihood approach". J Mol Evol. 17(6):368-76.

Filadelfi, A. M. and A. M. Castrucci. (1996). "Comparative aspects of the pineal/melatonin system of poikilothermic vertebrates". J Pineal Res. 20(4):175-86

Filadelfi, A. M., A. Vieira, F. M. Louzada. (2005). "Circadian rhythm of physiological color change in the amphibian *Bufo ictericus* under different photoperiods". Comp Biochem Physiol A Mol Integr Physiol. 142(3):370-5

Forsell, J., P. Ekström, I. N. Flamarique, B. Holmqvist. (2001). "Expression of pineal ultraviolet- and green-like opsins in the pineal organ and retina of teleosts". J. Exp. Boil 204(14): 2517-24.

Foster, R. G. and L. Kreitzman. (2014). "The rhythms of life: what your body clock means to you!" Exp Physiol. 99(4):599-606

Foster, R. G., and B. G. Soni. (1998). "Extraretinal photoreceptors and their regulation of temporal physiology". Reviews of Reproduction 3(3):145-50.

Fu, J. Z., Y. Z. Wang, X. M. Zeng, Z. Liu, Y. Zheng, J. D. McEachran. (2001). "Genetic diversity of eastern batrachuperus (Caudata: Hynobiidae)". Copeia. (4), 1100-1107

Gemmell, N. J. and J. Slate. (2006). "Heterozygote Advantage for Fecundity". PLoS ONE 1(1): e125

Goujon, M., H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, R. Lopez. (2010). "A new bioinformatics analysis tools framework at EMBL-EBI" Nucleic acids research. 38: W695-9

Graf, J. D. (1989a). "Genetic mapping in *Xenopus laevis*: eight linkage groups established". Genetics 123: 389–398.

Graf, J. D. (1989b). "Sex linkage of malic enzyme in *Xenopus laevis*". Experientia 45: 194–196.

Graf, J. D. (1993). In: O'Brien SJ (ed). Genetic Maps: Locus Maps of Complex Genomes. Cold Spring Harbor Laboratory Press: New York, 4330–4331.

Graf, J. D. and H. R. Kobel, (1991) Genetics of *Xenopus laevis*, in Methods in Cell Biology, B. Kay and H.B. Peng, Editors. Academic Press: San Diego, Ca.

Green, C. B. (2003). "Molecular control of *Xenopus* retinal circadian rhythms". J Neuroendocrinol 15:350-354

Guindon, S. and O. Gascuel. (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood". Syst Biol. 52(5):696-704.

Gurdon, J. B. (1960). "The developmental capacity of nuclei taken from differentiating endoderm cells of *Xenopus laevis*". J Embryol Exp Morphol. 8:505-26

Halford, S., M. S. Freedman, J. Bellingham, S. L. Inglis, S. Poopalasundaram, B. G. Soni, R.G. Foster, D. M. Hunt. (2001). "Characterization of a novel human opsin gene with wide tissue expression and identification of embedded and flanking genes on chromosome 1q43". Genomics. 1;72(2):203-8

Hanke, M. and M. Wink. (1994). "Direct DNA sequencing of PCR-amplified vector inserts following enzymatic degradation of primer and dNTPs". Biotechniques. 17(5):858-60

Hattar, S., R. J. Lucas, N. Mrosovsky, S. Thompson, R. H. Douglas, M. W. Hankins, J. Lem, M. Biel, F. Hofmann, R. G. Foster, K.W. Yau. (2003). "Melanopsin and rod-cone

photoreceptive systems account for all major accessory visual functions in mice". Nature. 3;424(6944):76-81.

Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, D. S. Rokhsar. (2010). "The genome of the Western clawed frog *Xenopus tropicalis*". Science 328: 633–636

Hirsch, N., L. B. Zimmerman, R. M. Grainger. (2002). "Xenopus, the next generation: *X. tropicalis* genetics and genomics". Dev Dyn. 225(4):422-33.

Hogan, B., R. Behringer, M. Gertsenstein, K. V. Nagy, S. Lunenfeld, A. Nagy, S. Lunenfeld, "Manipulating the mouse embryo: A laboratory manual (second edition)". 1994: Cold Spring Harbor Laboratory Press.

Hoque, B. and S. K. Saidapur. (1994). "Dynamics of oogenesis in the tropical anuran *Rana tigrina* (Amphibia; Ranidae) with special reference to vitellogenic cycles in wild-caught and captive frogs". J. Biosci. 19 339–352

Hubel, D., Origin of the Brain, 1988

Isorna, E., L. Besseau, G. Boeuf, Y. Desdevises, R. Vuilleumier, A. L. Alonso-Gómez, M. J. Delgado, J. Falcón. (2006). "Retinal, pineal and diencephalic expression of frog arylalkylamine N-acetyltransferase-1". Mol Cell Endocrinol. 27;252(1-2):11-8

Iyengar, B. (2013). "The melanocyte photosensory system in the human skin". SpringerPlus, 2:158

Katoh, K., K. Misawa, K. Kuma, T. Miyata. (2002). "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform". Nucleic Acids Res. 15;30(14):3059-66.

Kawano-Yamashita, E., M. Koyanagi, Y. Shichida, T. Oishi, S. Tamotsu, A. Terakita. (2011). "β-arrestin functionally regulates the non-bleaching pigment parapinopsin in lamprey pineal". PLoS One. 31;6(1):e16402.

Kawano-Yamashita, E., A. Terakita, M. Koyanagi, Y. Shichida, T. Oishi, S. Tamotsu. (2007) "Immunohistochemical characterization of a parapinopsin-containing photoreceptor cell involved in the ultraviolet/green discrimination in the pineal organ of the river lamprey *Lethenteron japonicum*" J. Exp. Boil **210** (21) 3821-9.

Kelleher, F. C., A. Rao, A. Maguire. (2014). "Circadian molecular clocks and cancer". Cancer Lett. 1;342(1):9-18.

Khokha, M. K., C. Chung, E. L. Bustamante, L. W. Gaw, K. A. Trott, J. Yeh, N. Lim, J. C. Lin, N. Taverner, E. Amaya, N. Papalopulu, J. C. Smith, A. M. Zorn, R. M. Harland, T. C. Grammer. (2002). "Techniques and probes for the study of *Xenopus tropicalis* development". Dev Dyn. 225(4):499-510

Kimmel, C.B., (1989). "Genetics and early development of the zebrafish". TIG, 4:284-288.

Klein, S. L., R. L. Strausberg, L. Wagner, J. Pontius, S. W. Clifton, P. Richardson. (2002). "Genetic and genomic tools for *Xenopus* research: The NIH *Xenopus* initiative". Dev. Dyn., 225: 384–391

Kojima, D., A. Terakita, T. Ishikawa, Y. Tsukahara, A. Maeda, Y. Shichida. (1997). "A novel Go-mediated phototransduction cascade in scallop visual cells". J Biol Chem. 12;272(37):22979-82.

Kojima, D., M. Torii, Y. Fukada, J. E. Dowling. (2008). "Differential expression of duplicated VAL-opsin genes in the developing zebrafish". J Neurochem. 104(5):1364-71.

Korf, H. W., C. Schomerus, J. H. Stehle. (1998). "The pineal organ, its hormone melatonin, and the photoneuroendocrine system". Adv. Anat. Embryol. Cell. Biol. 146, 1-100.

Kotpal, R. L. (2010). "Modern Text Book of Zoology: Vertebrates". Global Media Publications

Koyanagi, M. and A. Terakita, (2013), "Diversity of animal opsin-based pigments and their optogenetic potential", Biochim. Biophys. Acta 1837(5):710-6

Koyanagi, M., A. Terakita, K. Kubokawa, Y. Shichida. (2002). "Amphioxus homologs of Go-coupled rhodopsin and peropsin having 11-cis- and all-trans-retinals as their chromophores". FEBS Lett. 20;531(3):525-8.

Koyanagi, M., E. Kawano, Y. Kinugawa, T. Oishi, Y. Shichida, S. Tamotsu, A. Terakita. (2004). "Bistable UV pigment in the lamprey pineal". Proc. Nat. Acad. Sci USA **27**;101(17) 6687-91.

Koyanagi, M., K. Kubokawa, H. Tsukamoto, Y. Shichida, A. Terakita. (2005). "Cephalochordate melanopsin: Evolutionary linkage between invertebrate visual cells and vertebrate photosensitive retinal ganglion cells". Curr Biol, 15, 1065–1069

Koyanagi, M., K. Takano, H. Tsukamoto, K. Ohtsu, F. Tokunaga, A. Terakita. (2008). "Jellyfish vision starts with cAMP signaling mediated by opsin-G(s) cascade". Proc Natl Acad Sci U S A. 7;105(40):15576-80

Krieger, E., K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker, K. Karplus. (2009). "Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8". Proteins 77(9):114–122.

Kristiansen, K. (2004). "Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and

mutagenesis approaches to receptor structure and function". Pharmacol Ther. 103(1):21-80

Kruse, A. C., A. Manglik, B. K. Kobilka, W. I. Weis. (2013). "Applications of molecular replacement to G protein-coupled receptors". Acta Cryst. D69, 2287-2292

Kubo, Y., T. Takeuchi, K. Okano, T. Okano. (2010). "Cryptochrome Genes Are Highly Expressed in the Ovary of the African Clawed Frog, *Xenopus tropicalis*". PLoS ONE 5(2): e9273.

Kwok, P. and C. Xiangning. (2003). "Detection of single nucleotide polymorphisms". Curr. Issues Mol. Biol. 5: 43-60

Kwok, S., S. Y. Chang, J. J. Sninsky, A. Wang. (1994). "A guide to the design and use of mismatched and degenerate primers". PCR Methods Appl. 3(4):S39-47.

Kwoy, P. Y., C. Carlson, T. Yager, W. Ankener, D. A. Nickerson. (1994). "Comparative analysis of human DNA variation by fluorescence-base quenching of PCR products". Genomics 23 138-144,

Lagerkvist, A., J. Stewart, M. Lagerström-Fermér, U. Landegren. (1994). "Manifold sequencing: efficient processing of large sets of sequencing reactions". Proc Natl Acad Sci U S A. 15;91(6):2245-9.

Lagerström, M. C. and H. B. Schiöth. (2008). "Structural diversity of G protein-coupled receptors and significance for drug discovery". Nat Rev Drug Discov. 7(4):339-57

Lamb, T. D. (2009). "Evolution of vertebrate retinal photoreception". Phil. Trans. R. Soc. B 364 1531 2911-2924

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T.J. Gibson, D. G. Higgins. (2007). "ClustalW and ClustalX version 2". Bioinformatics 23(21): 2947-2948

- Leitner, T., D. Escanilla, C. Franzén, M. Uhlén, J. Albert. (1996). "Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis". PNAS 93 (20) 10864-10869
- Lesbarreres, D., D. S. Schmeller, C. R. Primmer, J. Merila." (2007). "Genetic variability predicts common frog (*Rana temporaria*) size at metamorphosis in the wild". Heredity 99, 41–46
- Li, J., P. C. Edwards, M. Burghammer, C. Villa, G. F. Schertler. (2004). "Structure of bovine rhodopsin in a trigonal crystal form". J Mol Biol. 5;343(5):1409-38
- Link, A. J., G. Skretas, E. M. Strauch, N. S. Chari, G. Georgiou. (2008). "Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in *Escherichia coli*". Protein Sci. 17(10):1857-63
- Lucas, P. W., N. J. Dominy, P. Riba-Hernandez, K. E. Stoner, N. Yamashita, E. Loría-Calderón, W. Petersen-Pereira, Y. Rojas-Durán, R. Salas-Pena, S. Solis-Madrigal, D. Osorio, B. W. Darvell. (2003). "Evolution and function of routine trichromatic vision in primates". Evolution. 57(11):2636-43.
- Luscombe, N. M., D. Greenbaum, M. Gerstein. (2001). "What is bioinformatics? A proposed definition and overview of the field". Methods Inf Med. 40(4):346-58
- MacCulloch, R. D., D. E. Upton, R. W. Murphy. (1996). "Trends in nuclear DNA content among amphibians and reptiles". Comparative Biochemistry and Physiology 113B: 601-605.
- Major, E., K. Rigo, T. Hague, A. Berces, S. Juhos. (2013). "HLA typing from 1000 genomes whole genome and whole exome illumina data". PLoS ONE 8(11):e78410
- Mano, H., D. Kojima, Y. Fukada. (1999). "Exo-rhodopsin: a novel rhodopsin expressed in the zebrafish pineal gland". Brain Res Mol Brain Res. 10;73(1-2):110-8

Masri, S. and P. Sassone-Corsi. (2013). "The circadian clock: a framework linking metabolism, epigenetics and neuronal function". Nat Rev Neurosci. 14, 69-75

Masri, S., M. Cervantes, P. Sassone-Corsi. (2013). "The circadian clock and cell cycle: interconnected biological circuits". Curr Opin Cell Biol. 25(6):730-4.

Matsuba, C. and J. Merilä. (2006). "Genome size variation in the common frog *Rana temporaria*". Hereditas. 143(2006):155-8

Matsuba, C., I. Miura, J. Merila. (2008). "Disentangling genetic vs. environmental causes of sex determination in the common frog, *Rana temporaria*". BMC Genet 9:3

Max, M., P. J. McKinnon, K. J. Seidenman, R. K. Barrett, M. L. Applebury, J. S. Takahashi, R. F. Margolskee. (1995). "Pineal opsin: a nonvisual opsin expressed in chick pineal". Science. 10;267(5203):1502-6.

Mayrand, P. E., L. B. Hoff, L. G. McBride, J. A. Bridgham, R. Cathcart, K. P. Corcoran, G. S. Golda, D. H. Keith, E. W. Lachenmeier, D. E. Madden, W. Mordan, M. W. Recknor, J. Shlgeura, C. Ting, N. W. Whfteley, J. S. Zlegie, M. N. Kronlck. (1990). "Automation of specific human gene detection". Clin Chem.;36(12):2063-71

McDiarmid, R. W. and R. Altig. (1999). "Tadpoles the biology of anuran larvae". University of Chicago Press

McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, R. Lopez. (2013). "Analysis tool web services from the EMBL-EBI". Nucleic acids research 41:W597-600

Mohammadi, M. T., R. Amini, Z. Jahanbakhsh, S. Shekarforoush. (2013). "Effects of atorvastatin on the hypertension-induced oxidative stress in the rat brain". Iran. Biomed. J. 17 (3): 152-157

- Moraes, M. N. C. M., M. O. Poletini, B. C. R. Ramos, L. H. R. G. Lima, A. M. L. Castrucci, (2013). "Effect of light on expression of clock genes in *Xenopus laevis* melanophores," Photochemistry and Photobiology, 90: 696–701
- Moraes, M. N., L. H. Lima, B. C. Ramos, M. D. Poletini, A. M. Castrucci. (2014). "Endothelin modulates the circadian expression of non-visual opsins." Gen Comp Endocrinol. pii: S0016-6480(14)00149-X
- Müller-Steinhardt, M., F. Schulte, H. Klüter, P. Bugert. (2009). "Optimized PCR with sequence specific primers (PCR-SSP) for fast and efficient determination of Interleukin-6 Promoter -597/-572/-174 Haplotypes". BMC Res Notes. 10;2:245.
- Murthy, S. K., L. M. DiFrancesco, R. T. Ogilvie, D. J. Demetrick. (2002). "Loss of heterozygosity associated with uniparental disomy in breast carcinoma". Mod Pathol. 15(12):1241-50.
- Musani, V., M. Sabol, D. Car, P. Ozretic, S. Oreskovic, D. Leovic, S. Levanat. (2012). "LOH of PTCH1 region in BCC and ovarian carcinoma: microsatellite vs. HRM analysis". Front Biosci (Elite Ed). 1;4:1049-57
- Okano, T., D. Kojima, Y. Fukada, Y. Shichida, T. Yoshizawa. (1992). "Primary structures of chicken cone visual pigments: vertebrate rhodopsins have evolved out of cone visual pigments". Proc Natl Acad Sci U S A. 1;89(13):5932-6.
- Okano, T., T. Yoshizawa, Y. Fukada. (1994). "Pinopsin is a chicken pineal photoreceptive molecule". Nature. 3;372(6501):94-7
- Olerup, O. and H. Zetterquist. (1991). "HLA-DRB1*01 subtyping by allele-specific PCR amplification: a sensitive, specific and rapid technique". Tissue Antigens. 37(5):197-204
- Olerup, O., H. Zetterquist, E. Möller. (1991). "Identification of HLA class II polymorphism with molecular biology techniques". Lakartidningen. 88(40):3275-7

- Pakkasmaa, S., J. Merila and R. B. O'Hara. (2003). "Genetic and maternal effect influences on viability of common frog tadpoles under different environmental conditions". Heredity 91, 117–124
- Palo, J. U., D. Lesbarrères, D. S. Schmeller, C. R. Primmer, J. Merilä. (2004). "Microsatellite marker data suggest sex-biased dispersal in the common frog *Rana temporaria*". Mol Ecol. 13(9):2865-9.
- Panda, S., I. Provencio, D. C. Tu, S. S. Pires, M. D. Rollag, A. M. Castrucci, M. T. Pletcher, T.K. Sato, T. Wiltshire, M. Andahazy, S. A. Kay, R. N. Van Gelder, J. B. Hogenesch. (2003). "Melanopsin is required for non-image-forming photic responses in blind mice". Science. 25;301(5632):525-7.
- Papp, A. C., J. K. Pinsonneault, G. Cooke, W. Sadée. (2003). "Single nucleotide polymorphism genotyping using allele-specific PCR and fluorescence melting curves". BioTechniques 34:1068-1072
- Patel, A. B., E. Crocker, M. Eilers, A. Hirshfeld, M. Sheves, S. O. Smith. (2004). "Coupling of retinal isomerization to the activation of rhodopsin". Proc Natl Acad Sci U S A. 6;101(27):10048-53.
- Peirson, S. N., S. Halford and R. G. Foster. (2009). "The evolution of irradiance detection: melanopsin and the non-visual opsins". Phil. Trans. R. Soc. B 364
- Plachetzki, D. C., B. M. Degnan, T. H. Oakley. (2007). "The origins of novel protein interactions during animal opsin evolution". PLoS One. 17;2(10):e1054.
- Poeggeler, B., S. Saarela, R. J. Reiter, D. X. Tan, L. D. Chen, L. C. Manchester, L. R. Barlow-Walden. (1994). "Melatonin—a highly potent endogenous radical scavenger and electron donor: new aspects of the oxidation chemistry of this indole accessed in vitro". Ann N Y Acad Sci. 738:419–420

- Pope, A., M. Eilers, P. J. Reeves, S. O. Smith. (2014). "Amino acid conservation and interactions in rhodopsin: probing receptor activation by NMR spectroscopy." Biochim Biophys Acta. 1837(5):683-93
- Provencio, I., G. Jiang, W. J. De Grip, W. P. Hayes, M. D. Rollag. (1998). "Melanopsin: An opsin in melanophores, brain, and eye". Proc Natl Acad Sci USA. 6;95(1):340-5
- Provencio, I., I. R. Rodriguez, G. Jiang, W. P. Hayes, E. F. Moreira, M. D. Rollag. (2000). "A novel human opsin in the inner retina". J. Neurosci. 20, 6000-605.
- Radu, R. A., J. Hu, J. Peng, D. Bok, M. L. Mata, G. H. Travis.(2008). "Retinal pigment epithelium-retinal G protein receptor-opsin mediates light-dependent translocation of all-trans-retinyl esters for synthesis of visual chromophore in retinal pigment epithelial cells". J Biol Chem. 11;283(28):19730-8.
- Reed, B. T. (2005). "Guidance on the housing and care of the African clawed frog *Xenopus laevis*". Animal research department RSPCA report.
- Refinetti, R., and M. Menaker. (1992). "The circadian rhythm of body temperature". Physiology and Behavior, 51, 613–637
- Richter, A., F. Grieu, A. Carrello, B. Amanuel, K. Namdarian, A. Rynska, A. Lucas, V. Michael, A. Bell, S .B. Fox, C. A. Hewitt, H. Do, G. A. McArthur, S. Q. Wong, A. Dobrovic, B. Iacopetta. (2013). "A multisite blinded study for the detection of BRAF mutations in formalin-fixed, paraffin-embedded malignant melanoma". Sci Rep. 3:1659.
- Ridge, K. D., N. G. Abdulaev, M. Sousa, K. Palczewski. (2003). "Phototransduction: crystal clear". Trends Biochem Sci. 28(9):479-87.
- Roelants, K., D. J. Gower, M. Wilkinson, S. P. Loader, S. D. Biju, K. Guillaume, L. Moriau, F. Bossuyt. (2007). "Global patterns of diversification in the history of modern amphibians". PNAS. 104 (3) 887-892

- Rosenbaum, D. M., S. G. Rasmussen, B. K. Kobilka. (2009) "The structure and function of G-protein-coupled receptors". Nature. 21;459(7245):356-63
- Saba, N. and N. K.Tripathi. (2012). "Meiotic Chromosomes and Karyotype of Bufo viridis (Laurenti,1768) from Jammu and Kashmir". Bull. Environ. Pharmacol. Life Sci. 1(6): 21-25.
- Saitou, N. and M. Nei. (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees". Mol Biol Evol. 4(4):406-25.
- Sakai, K., Y. Imamoto, C. Y. Su, H. Tsukamoto, T. Yamashita, A. Terakita, K. W. Yau, Y. Shichida. (2012). "Photochemical nature of parietopsin". Biochemistry. 6;51(9):1933-41
- Sambrook J., Fritsch E. and Maniatis T. (1989). Molecular Cloning, A Laboratory Manual (Secound edition). Volumes I, II and III. Cold spring Harbour Laboratory Press, USA.
- Sato, K., T. Yamashita, H. Ohuchi, Y. Shichida. (2011) "Vertebrate ancient long opsin has molecular properties intermediate between those of vertebrate and invertebrate visual pigments". Biochemistry. 50, 10484–10490
- Savage, D. A., J. P. Tang, N. A. Wood, J. Evans, J. L. Bidwell, J. L. Wee, A. A. Oei, K. M. Hui. (1996). "A rapid HLA-DRB1*04 subtyping method using PCR and DNA heteroduplex generators". Tissue Antigens. 47(4):284-92
- Schäffer, A. A., L. Aravind, T. L. Madden, S. Shavirin, J. L. Spouge, Y. I. Wolf, E. V. Koonin, S. F. Altschul. (2001), "Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements", Nucleic Acids Res. 29:2994-3005
- Schmeller, D. S., J. Schregel, M. Veith. (2007). "The importance of heterozygosity in a frog's life". Naturwissenschaften. 2007 May;94(5):360-6.

Schulz HL1, Goetz T, Kaschkoetoe J, Weber (2004). "The Retinome - defining a reference transcriptome of the adult mammalian retina/retinal pigment epithelium" 2004 July 29;5(1):50

Sellis, D., B. J. Callahan, D. A. Petrov, P. W. Messer. (2011). "Heterozygote advantage as a natural consequence of adaptation in diploids". PNAS. 108 (51) 20666-20671

Shichida, Y. and T. Matsuyama. (2009). "Evolution of opsins and phototransduction". Phil. Trans. R. Soc. B 364, 2881–2895

Shichida, Y. and T. Yamashita. (2003). "Diversity of visual pigments from the viewpoint of G protein activation—comparison with other G protein-coupled receptors." Photochem. Photobiol. Sci. 2, 1237–1246

Shoichet, B. K. and B. K. Kobilka. (2012). "Structure-based drug screening for G-protein-coupled receptors." Trends Pharmacol Sci. 33(5):268-72

Simsek, M., M. O. M. Tanira, K. A. Al-Baloushi, H. S. Al-Barwani, K. M. Lawatia, R. A.L. Bayoumi. (2001). "A precaution in the detection of heterozygotes by sequencing: comparison of automated DNA sequencing and PCR-restriction fragment length polymorphism methods". Clinical Chemistry. 47, 134-137

Skretas, G. and G. Georgiou. (2009). "Genetic analysis of G protein-coupled receptor expression in *Escherichia coli*: inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor." Biotechnol Bioeng. 1;102(2):357-67

Skretas, G., T. Makino, N. Varadarajan, M. Pogson, G. Georgiou. (2012). "Multi-copy genes that enhance the yield of mammalian G protein-coupled receptors in *Escherichia coli*". Metab Eng. 14(5):591-602

Snelson, C. D., K. Santhakumar, M. E. Halpern, J. T. Gamse. (2008). "Tbx2b is required for the development of the parapineal organ". Development 135(9):1693-702

Soni, B. G. and R. G. Foster. (1997). "A novel and ancient vertebrate opsin". FEBS Lett. Apr 14;406(3):279-83

Srivatsan, A., Y. Han, J. Peng, A. K. Tehranchi, R. Gibbs, J. D. Wang, R. Chen. (2008). "High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies". PLoS Genet. 1;4(8):e1000139.

Stanley, K. A. (2008). "Semi-volatile Organic Compounds and Developing Organisms: Accumulation in California Mountain Tadpoles in the Field and Fish Embryo Exposures in the Laboratory". ProQuest LLC

Stehle, J. H., C. von Gall, H. W. Korf. (2001). "Analysis of cell signalling in the rodent pineal gland deciphers regulators of dynamic transcription in neural/endocrine cells". Eur. J. Neurosci. 14, 1-9.

Stella, S. L. Jr., A. Vila, A. Y. Hung, M. E. Rome, U. Huynh, M. Sheng, H. J. Kreienkamp, N. C. Brecha. (2012). "Association of shank 1A scaffolding protein with cone photoreceptor terminals in the mammalian retina". PLoS One. 7(9):e43463

Stuart, S. N., J. S. Chanson, N. A. Cox, B. E. Young, A. S. Rodrigues, D. L. Fischman, R. W. Waller. (2004). "Status and trends of amphibian declines and extinctions worldwide". Science 306 (5702), 1783-1786

Su, C. Y., D. G. Luo, A. Terakita, Y. Shichida, H. W. Liao, M. A. Kazmi, T. P. Sakmar, K. W. Yau. (2006). "Parietal-eye phototransduction components and their potential evolutionary implications". Science. 17;311(5767):1617-21

Suga, H., V. Schmid, W. J. Gehring. (2008). "Evolution and functional diversity of jellyfish opsins". Curr Biol. 8;18(1):51-5.

Takahama, H. (1993). "Evidence for a frontal-organ homologue in the pineal complex of the salamander, *Hynobius dunni*." Cell Tissue Res. 272, 575-578.

Takahashi, J. S., F. W. Turek, R. Y. Moore. (2001). "Circadian Clocks", Vol 12 Springer, Medical – 116-117

Talaat, R. M., M. F. Dondeti, S. Z. El-Shenawy, O. A. Khamiss. (2013). "Transforming Growth Factor- β 1 Gene Polymorphism (T29C) in Egyptian Patients with Hepatitis B Virus Infection: A Preliminary Study". Hepat Res Treat. 2013:293274.

Taniguchi, Y., O. Hisatomi, M. Yoshida, F. Tokunaga. (2001). "Pinopsin expressed in the retinal photoreceptors of a diurnal gecko". FEBS Lett. 11;496(2-3):69-74.

Tarttelin, E. E., J. Bellingham, L. C. Bibb, R. G. Foster, M. W. Hankins, K. Gregory-Evans, C. Y. Gregory-Evans, D. J. Wells, R. J. Lucas. (2003). "Expression of opsin genes early in ocular development of humans and mice". Exp Eye Res. 76(3):393-6.

Terakita, A. (2005). "The opsins." Genome Biology, 6:213

Terakita, A., H. Tsukamoto, M. Koyanagi, M. Sugahara, T. Yamashita, Y. Shichida. (2008) "Expression and comparative characterization of Gq-coupled invertebrate visual pigments and melanopsin". J. Neurochem. 105, 883–890.

Terakita, A., M. Koyanagi, H. Tsukamoto, T. Yamashita, T. Miyata, Y. Shichida. (2004) "Counterion displacement in the molecular evolution of the rhodopsin family". Nat. Struct Mol. Biol 11(3) 284-9.

The Nobel Prize in Chemistry 2012, Robert J. Lefkowitz and Brian K. Kobilka "for studies of G-protein–coupled receptors". PRESSMEDDELANDE Press release 2012

Thompson, J. D., D. G. Higgins, T. J. Gibson. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". Nucleic Acids Res., 22, 4673-4680

Tinsley, R. and H.R. Kobel, eds. The Biology of Xenopus. . 1996, Clarendon Press: Oxford.

Tosini, G., C. Bertolucci, A. Foa. (2001). "The circadian system of reptiles: a multioscillatory and multiphotoreceptive system". Physiol. Behav. 4, 461-471.

Tsukamoto, H., A. Terakita, Y. Shichida. (2010). "A pivot between helices V and VI near the retinal-binding site is necessary for activation in rhodopsins." J Biol Chem. 285:10, 7351–7357.

Tsukamoto, H. and A. Terakita. (2010). "Diversity and functional properties of bistable pigments". Photochemical & Photobiological Sciences 9(11) 1435-43.

Tsukamoto, H., D. L. Farrens, M. Koyanagi, A. Terakita. (2009) "The magnitude of the light-induced conformational change in different rhodopsins correlates with their ability to activate G proteins". J Biol Chem 284(31) 20676-83.

Tuttle, T. R., E. R. Hugo, W. S. Tong, N. Ben-Jonathan. (2014). "Placental Lactogen Is Expressed but Is Not Translated into Protein in Breast Cancer". PLoS ONE 9(1):e87325

Ukoskit, K. (2001). "Detection of single nucleotide polymorphisms in cattle genes using automated DNA sequencing and dideoxy fingerprinting". Thammasat Int. J.'Sc. Tech., 6(3)

Ung, C. Y. and A. C. Molteno . (2004). "An enigmatic eye: the histology of the tuatara pineal complex". Clinical & Experimental Ophthalmology. 32(6):614-8.

Vallone, D., K. Lahiri, T. Dickmeis, S. F. Nicholas. (2005) "Zebrafish cell clocks feel the heat and see the light"! Zebrafish 2(3): 171-187.

Venselaar, H., T. A. Te Beek, R. K. Kuipers, M. L. Hekkelman, G. Vriend. (2010). "Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces." BMC Bioinformatics. 11:548

Vieites, D. R., S. Nieto-Román, M. Barluenga, A. Palanca, M. Vences, A. Meyer. (2004). "Post-mating clutch piracy in an amphibian". Nature. 16;431(7006):305-8.

Vígh, B., M.J. Manzano, A. Zádori, C.L. Frank, A. Lukáts, P. Röhlich, A. Szél and C. Dávid. (2002). "Nonvisual photoreceptors of the deep brain, pineal organs and retina". Histol Histopathol. 17: 555-590

Vígh-Teichmann, I. and B. Vígh. (1990). "Opsin immunocytochemical characterization of different types of photoreceptors in the frog pineal organ". J. Pineal Res. 8, 323-333.

von Schantz, M., I. Provencio, R. G. Foster. (2000). "Recent developments in circadian photoreception: more than meets the eye". IOVS. 41:1605-1607

Wada, S., E. Kawano-Yamashita, M. Koyanagi, A. Terakita. (2012). "Expression of UV-sensitive parapinopsin in the Iguana parietal eyes and its implication in UV-sensitivity in vertebrate pineal-related organs". PLoS ONE 7(6): e39003.

Wada, Y., T. Okano, A. Adachi, S. Ebihara, Y. Fukada. (1998). "Identification of rhodopsin in the pigeon deep brain". FEBS Lett. 6;424(1-2):53-6

Wagner, E. and J. Lykke-Andersen. (2002). "mRNA surveillance: the perfect persist." J Cell Sci 115:3033-3038

Wake, D. B. and V. T. Vredenburg. (2008). "Colloquium paper: are we in the midst of the sixth mass extinction? A view from the world of amphibians." Proc Natl Acad Sci U S A. 12;105 Suppl 1:11466-73

Welsh, K. and M. Bunce. (1999). "Molecular typing for the MHC with PCR-SSP". Rev Immunogenet. 1(2):157-76.

Woodhouse, M., D. Burkart-Wac, and L. Comai. (2009). "Polyploidy". Nature Education 2(1).

Wu, C. Z., X. J. Ni, S. L. Zheng, Y. R. Yang, P. Xia, Y. J. Zeng, B. C. Chen. (2009). "A fast SSP-PCR method for genotyping the ATP-binding cassette subfamily B member 1

gene C3435T and G2677T polymorphisms in Chinese transplant recipients." Tumori. 95: 338-342

Yamashita, T., K. Ono, H. Ohuchi, A. Yumoto, H. Gotoh, S. Tomonari, K. Sakai, H. Fujita, Y. Imamoto, S. Noji, K. Nakamura, Y. Shichida. (2014). "Evolution of mammalian Opn5 as a specialized UV-absorbing pigment by a single amino acid mutation". J Biol Chem. 14;289(7):3991-4000

Yau, K. W. and R. C. Hardie RC. (2009). "Phototransduction motifs and variations". Cell. 16;139(2):246-64

Yokoyama, S. (2000). "Molecular evolution of vertebrate visual pigments". Prog Retin Eye Res. 19(4):385-419.

Zavodna, K., T. Krivulcik, M. G. Bujalkova, T. Slamka, D. Martinicky, D. Ilencikova, Z. Bartosova. (2009). "Partial loss of heterozygosity events at the mutated gene in tumors from MLH1/MSH2 large genomic rearrangement carriers". BMC Cancer. 20;9:405.

Zelinski, E. L., S. H. Deibel, R. J. McDonald. (2014). "The trouble with circadian clock dysfunction: Multiple deleterious effects on the brain and body". Neurosci. Biobehav. Rev. 40:80-101

Zeng, X. M., J.Z. Fu, L.Q. Chen, Y.Z. Tian, X.H. Chen. (2006). "Cryptic species and systematics of hynobiid salamanders of the Liua–Pseudohynobius complex: molecular and phylogenetic perspectives". Biochem. Syst. Ecol., 34 467–477

Zhu, H., C. B. Green. (2001). "Three cryptochromes are rhythmically expressed in *Xenopus laevis* retinal photoreceptors". Mol Vis. 29;7:210-5.

Zhukovsky, E. A., D. D. Oprian. (1989). "Effect of carboxylic acid side chains on the absorption maximum of visual pigments". Science. 17; 246 (4932):928-30

Websites:

<http://animaldiversity.ummz.umich.edu/accounts/Amphibia/>

<http://faculty.virginia.edu/xtropicalis/overview/intro.html>

<http://genome.jgi-psf.org/Xentr4/Xentr4.info.html>

<http://genome-euro.ucsc.edu/cgi-bin/hgGateway>

http://gqinnovationcenter.com/documents/sequencing/UserGuide_SangerSequencing_en.pdf

<http://tropicalis.berkeley.edu/home/index.html>

<http://www.genecodes.com/>

<http://www.genomesize.com/results.php?page=1>

<http://www.ncbi.nlm.nih.gov/nuccore>

http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2012/press.html

<http://www.xenbase.org/anatomy/showanatomy.do?method=displayAnatomySummary&anatomyId=11>

<http://xgc.nci.nih.gov/Reagents/MGCCloneList> xenopus gene collection

http://upload.wikimedia.org/wikipedia/commons/c/c9/GPCR_cycle.jpg

www.xenbase.org

Appendix

Appendix 1

Xenopus tropicalis Parapinopsin (NM_213665)

```
1  tgttgccctg acaatttcgg tgactaaaa gtctcagaac caaatggccg
acgaggctct
61  tctcccacca atgatgaatg tgaccaatga agagatgcac cctgggaagg
ttctgatgcc
121 cgggattggg tacaccattc tggctttaat aatggccgta ttctgtgccg
ccgctctctt
181 tctcaacgtc acggtgattg tggtgacttt caaatatcgc caactgcgcc
acccaatcaa
241 ctactcgctg gtcaacctgg ccatcgccga ccttggcgtc accgtgctgg
gcggggcctt
301 gacagtggag acaaatgccg tgggctactt taacctgggg agagtgggct
gcgtcattga
361 gggtttcgcc gtggcatttt ttggcatcgc ggctttgtgc accattgcag
taatagcgtt
421 ggaccgagtg ttcgtgggtg gcaagccaat gggcacactc acattcacco
ccaaacaagc
481 actggccggc attgctgctt cctggatctg gtcgctcata tggaatacgc
ccccgctggt
541 cggctggggg agctacgagc tggaaggggt gatgacatcc tgcgccccca
actggtacag
601 cgccgaccca gtcaacatgt cctacatcgt ctgctacttc tccttctgct
ttgccatccc
661 attcctcatc atcgtggggg cctacgggta cctgatgtgg actctgcggc
aggttgccaa
721 gttgggtgta gcggaaggag gtacaacgag taaagccgag gtgcaagtct
cccgatatgg
781 gattgtcatg atcttggctt tcctgggtctg ttggctcccg tacgccgcct
ttgcatgac
841 ggttgtggca aatcccggaa tgcacattga cccattata gccaccgtac
ccatgtacct
901 gaccaaacc agcaccgtct ataatccaat tatctacatt ttcatgaaca
agcagttcca
961 agaattgtgc attccctttc tgttctgcgg gaggaatcct tgggctgctg
aaaagtcaag
1021 ttctatggaa acttctatca gcgtaacgag tggaacgcct acgaaacgtg
gccaagtggc
1081 tccggcgtaa cgtgctttta tctccagtat aaccggc
```

Appendix 2

Synthesized Parapinopsin Sequence

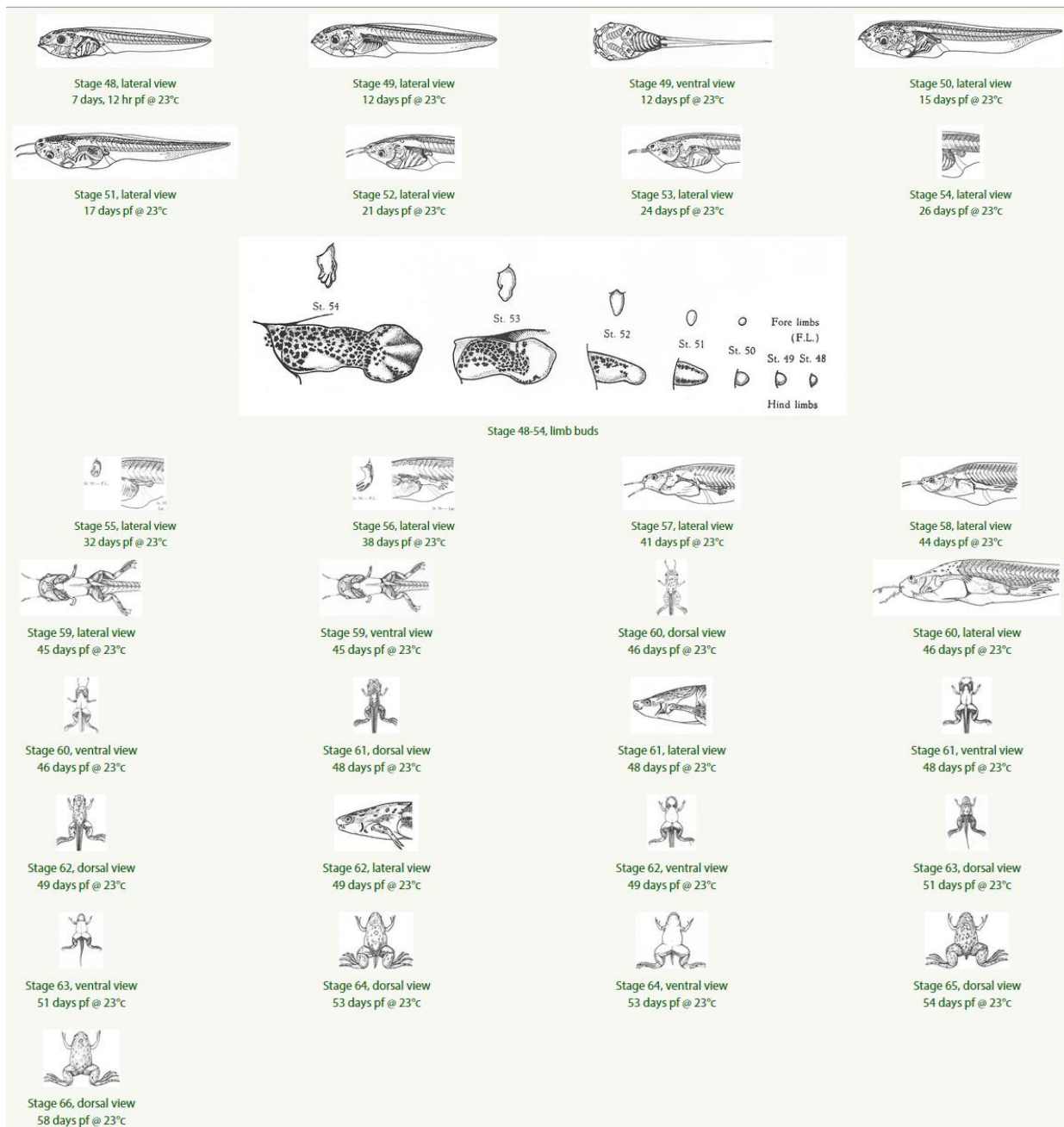
Cloned ParaP Sequence

GCATGCATGGCCGACGAGGCCCTGTTACCGCCTATGATGAACGTGACCAATGAGGAGATGCAT
CCGGGCAAAGTGCTGATGCCGCGCATTGGCTATACCATCCTGGCATTAAATCATGGCCGTTTTTT
GCGCCGACGCCCTGTTCTGAATGTGACCGTTATTGTGGTGACCTTTAAATATCGCCAGCTGCG
CCATCCGATCAACTACAGCCTGGTGAATTTAGCCATTGCCGATCTGGGCGTGACCGTTCTGGGT
GGCGCACTGACCGTGGAACCAATGCCGTGGGCTACTTTAATTTAGGCCGCGTGGGCTGCGTG
ATTGAGGGTTTCGCCGTTGCCTTTTTCGGCATTGCCGCCCTGTGCACCATTGCAGTGATCGCACT
GGACCGCGTGTGTTGTGGTTTGTAACCGATGGGCACCCTGACCTTCACCCCGAAACAGGCACTG
GCAGGTATTGCAGCCAGCTGGATCTGGAGCCTGATTTGGAACACCCCGCCGCTGTTTCGGTTGG
GGCAGCTATGAACTGGAAGGCGTTATGACCAGCTGCGCCCCGAATTGGTATAGCGCAGACCCG
GTGAACATGAGCTATATTGTTTGCTATTTTAGCTTTTGCTTTGCCATCCCGTTCTTAATCATCGTG
GGCAGCTACGGCTATCTGATGTGGACCCTGCGCCAGGTGGCCAAACTGGGTGTGGCCGAAGG
CGGCACCACAGTAAAGCCGAAGTGCAGGTAGCCGCATGGTGATCGTGATGATCCTGGCATT
TCTGGTTTGCTGGCTGCCGTATGCCGCCCTTTGCCATGACCGTGGTGGCCAATCCGGGGATGCAC
ATTGACCCGATTATCGCAACCGTGCCGATGTATCTGACCAAAACCAGCACCGTTTATAACCCTAT
TATCTATATTTTTATGAACAAGCAGTTTCAGGAGTGCGTGATCCCTTTCTGTTTTGTGGCCGTA
ACCCGTGGGCGAGCCGAAAAAAGCAGCAGCATGGAAACCAGTATCAGCGTGACCAGCGGTACC
CCGACCAAACGTGGCCAGGTGGCCCCGGCACCTGCAGG

Appendix 3



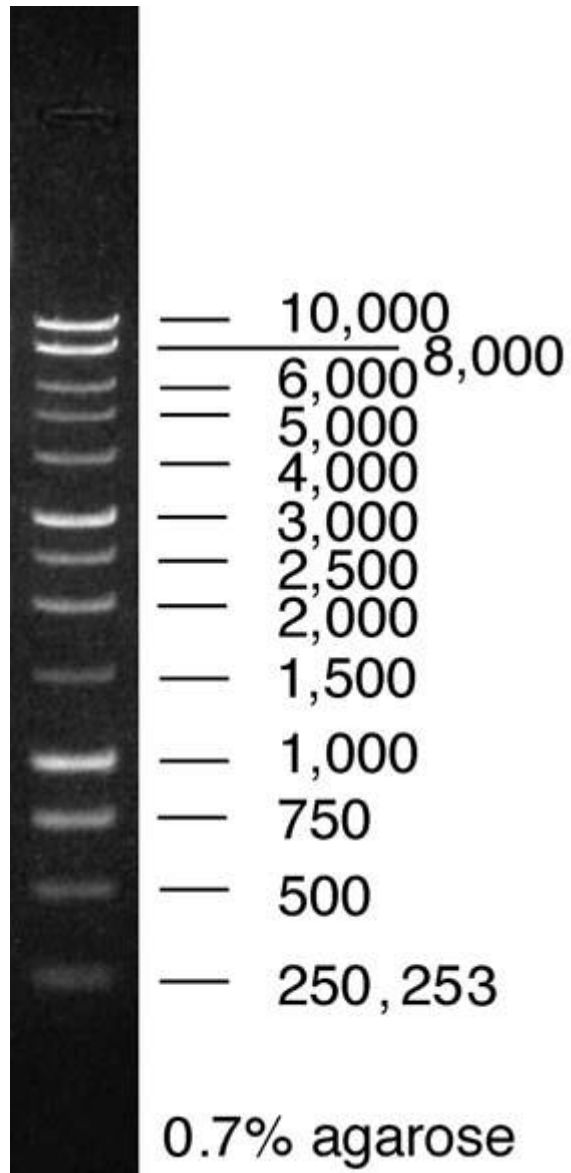




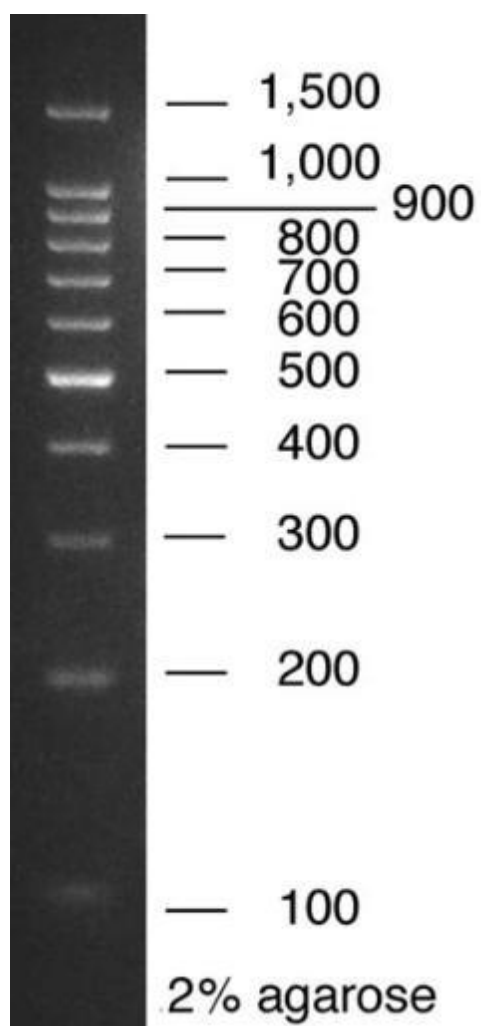
Appendix 4

DNA ladders used

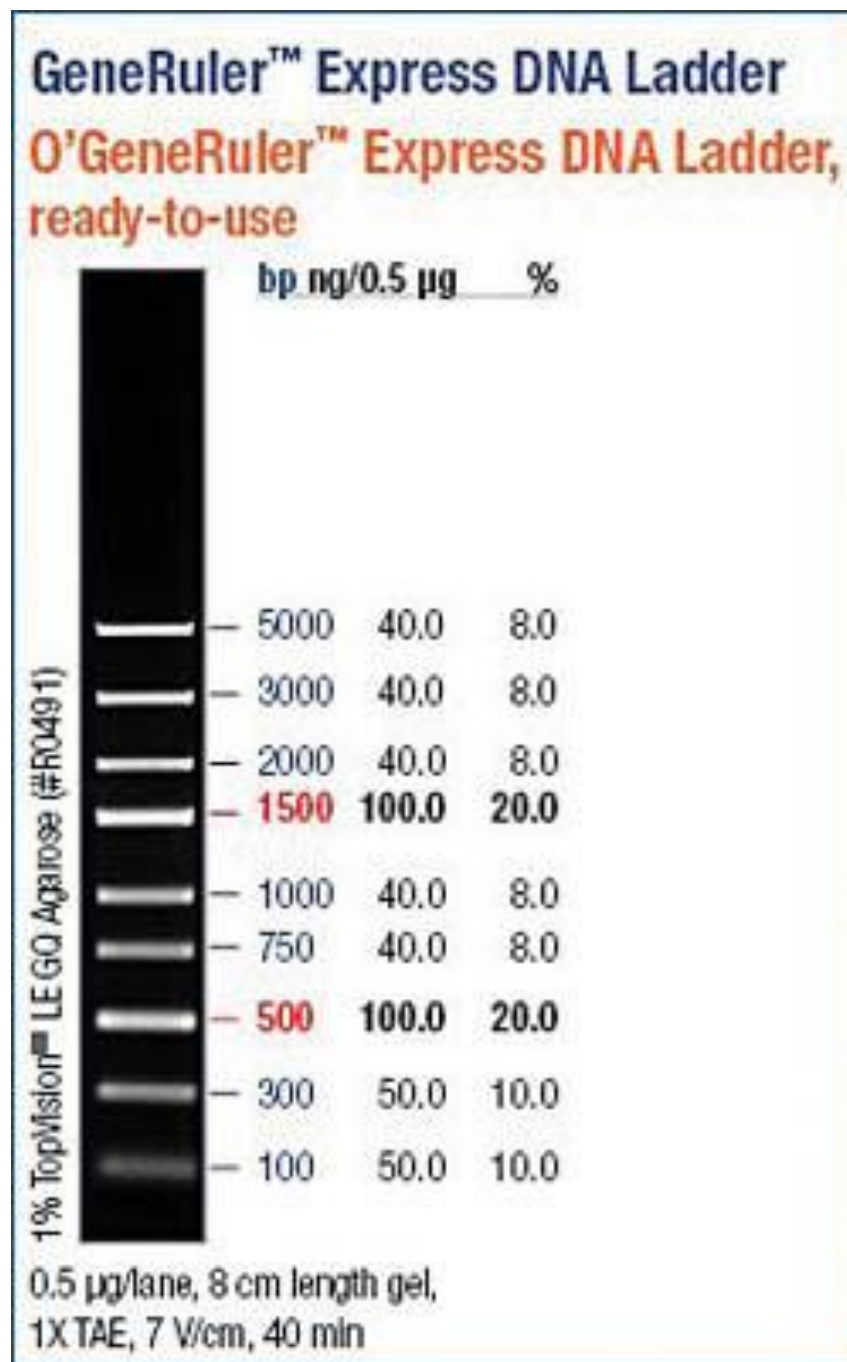
I. GeneRuler 1kb DNA ladder



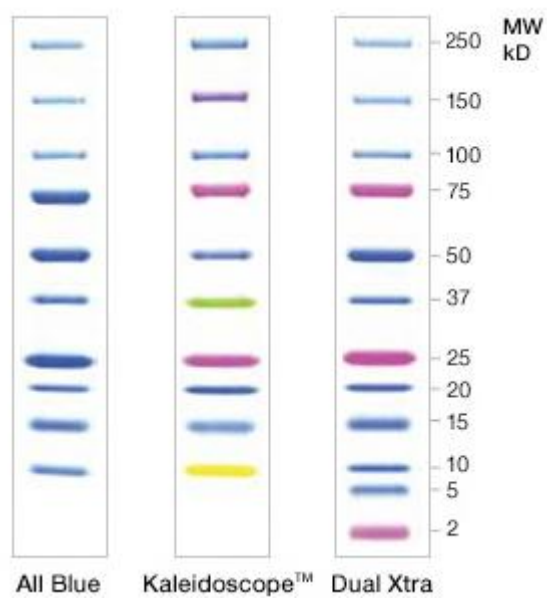
II. GeneRuler 100bp DNA Ladder



III. GeneRuler Express DNA ladder



IV. BioRad Protein Marker



Appendix 5

>*X. tropicalus* Parapinopsin (NM_213665.1)

MADEALLPPMMNVTNEEMHPGKVLMPRIGYTILALIMAVFCAAALFLNVTVIVVTFKYRQLRHPIN
YSLVNLAIADLGVTVLGGALTVETNAVGYFNLGRVGCVIEGFAVAFFGIAALCTIAVIALDRVFVCKP
MGTLTFTPQKALAGIAASWIWSLIWNTPLFGWGSYELEGVMTSCAPNWYSADPVNMSYIVCYFS
FCFAIPFLIIVGSYGylmWTLRQVAKLGVAEGGTSSKAQVSRMVIVMILAFLVCWLPYAAFAMT
VVANPGMHIDPIIATVPMYLTSTVYNPIIYIFMNKQFQECVIPFLFCGRNPWAAEKSSSMETSISV
TSGTPTKRGQVAPA

>Lamprey Parapinopsin (BAD13381.1)

MENLTSLDLLPNGEVPLMPRYGFTILAVIMAVFTIASLVLNSTVVIVTLRHRQLRHPLNFSVLNLAVA
DLGVTVFGASLVVETNAVGYFNLGRVGCVIEGFAVAFFGIAALCTIAVIAVDRFVVCKPLGTLMTFR
RHALLGIAWAWLWSFVWNTPLFGWGSYELEGVRTSCAPDWYSRDPANVSYITSYFAFCFAIPFLV
IVVAYGRLMWTLHQVAKLGMGESGSTAKAEAQVSRMVVVMVVAFLVCWLPYALFAMIVVTKPD
VYIDPVIATLPMYLTSTVYNPIIYIFMNRQFRDCAVPFLCGRNPWAEPSSESATAASTSATSVTLA
SAPGQVSPS

>Cat Fish Parapinopsin (NM_001200073.1)

MASIILINFSETDTLHLGSVNDHIMPRIGYTILSIIMALSSTFGIILNMVVIIVTVRYKQLRQPLNYALVN
LAVADLGCPVFGGLLTAVTNAMGYFSLGRVGCVLEGFAVAFFGIAGLCSVAVIAVDYRMVVCRLG
AVMFQTKHALAGVVFWSWVWSFIWNTPLFGWGSYQLEGVMTSCAPNWYRRDPVNVSYILCYFM
LCFALPFATIIFSYMHLHTLWQVAKLQVADSGSTAKVEVQVARMVVMVMAFLLTWLPYAAFALT
VIIDSNIYINPVIGTIPAYLAKSSTVFNPPIIYIFMNRQFRDYALPCLLCGKNPWAAKEGRSDTNTLT
VSKNTSVSPL

>Rainbow trout Parapinopsin (NM_001124225.1)

MDHQQLLPNLHGNISSPGSVSEALLSRTGFTILAVIIGVFSVSGVCMNVLVIMVTMRHRKLRQPLN
YALVNLAVADLGCALFGGLPTMVTNAMGYFSMGRLGCVLEGFAVAFFGIAGLCSVAVIAVDYRVV
VCRPMGAVMFQTRHAVGGVLSWVWSFLWNTPLFGWGSFELEGVRTSCSPNWYSREPGNMS
YIILYFLLCFAIPFSIIMVSYARILFTLHQVSKLKVLEGNSTTRVEIQVVRMVVVMVMAFLLSWLPYAAF
ALSVILDPSLHINPLIATVPMYLAKSSTVYNPIIYVFMNRQFRDCAVPFLCGLNPWASEPVGSEADT
ALSSVSKNPRVSPQ

>Zebrafish Parapinopsinb (NM_001005312.1)

MDFFPEFLNDSSTHYDGPGLKVPLPRAGFITLSLLMAVFSITSVVLNATVIVVTLRHKQLRQPLNFAL
VNLAVADLGTTLTGSVPSVVTNAVGYIMGRIGCVLEGFCVAFFGISALCTVALIAVERLFVVCRLPLG
SITFQCRHAAGLLSCWLWSLIWNTPLLLGWGSYQLEGAGTSCGPHWQSRELRDVSYIICYFSVCF
AVPFAIILVSYSWLLYTLRQVLLEMVNNLGTFGMLFSAYYDLGHIFKHKHCNAALTLVFFVVEYH
WLPYAALALTVVSKPEVQLAVLVKVLPIYMAKSSTVYNPLIYIMNKQVSRTS

>Green Anole Parapinopsin (NM_001293131.1)

MDSLDTNTLSPNASTVRVVLMPRIGYTIIAIIMATSCTLSVILNTAVIAITIKYRQLRQPINYSLVNLAI
DLGAALLGGSLNVETNAVGYYNLGRVGCVTEGFAMAFFGIVALCTIAVIAVDRAIVIAKPMGTITFTT
RKAMIGVAVSWIWSLVWNTPLFGWGGYQMEGVMTSCAPDWANSDPINVSIIICYFLCFTIPFI
TILASYGYLIWTLRQVAKVGLAQRGSTTKAEQVSRMVIVMVMAFLICWLPYATFALVVVGNPQIYI
NPIIATIPMYMAKSSTFYNPPIIYIFMNKQFRDCLVRCLLCGRNPCASEQTDEDDLEVSTIAPAPSSRRG
KVAPV

>Common iguana Parapinopsin (AB626969.1)

MDFLDTSTLSMNATTVRVVLMPRIGYTIIAIIMATSCTLSVILNTTVIAVTIKYRQLRQPINYSLVNLAI
ADLGAALLGGSLNVETNAVGYYNLGRVGCVTEGFAMAFFGIVALCTIAVIAVDRAIVIAKPMGTITFT
TTRKAMIGVAISWIWSLVWNTPLFGWGGYQMEGVMTSCAPDWYNSDPINVSIVCYFLCFTIPFI
FVTILVSYGYLLWTLRQVAKFGVTQRGSTNKAEQVSRMVIVMVMAFLICWLPYATFALVVVGNP
QIYINPIIATVPMYMAKSSTFYNPPIIYIFMNKQFRDCLVRCLLCGRNPCASDQAEEDDLEVSTIAPAPS
SKRGKVTM

>Iguana Pinopsin (AB626971.1)

MRAANEPNNGTPGPFEGPQWPYLASRIYTS LAVLMGLVVISAAIVNALVIAVSIQYKKLRSPLNYIL
VNLAIADLLVTSFGSTISFANNIYGFFVFGQAACKFEGFMVSLTGIVGLWSLAILALERYLVVCKPAGD
FRFQQRHALIGCVFTWAWSLAWTLPLFGWSSYVPEGLKTSCGPNWYTGGSSNNSYITTLFVTCFA
TPLGMIVFSYANLLTLRAVAAQQKESRTTRRAEREVTRMVIAMVLAFLVCWLPYATFAMVVATNK
DLVIQPALASLPSYFSKTATVYNPIIYVFMNKQFRSCLLYTMRCGRRPRDERETPPPRISVPRGPSSAL
EGSRNKVTPSASEGSGNDTRPS

>Cherry Salmon ParapinopsinB (AB772413.1)

MMQLPASLPNASSYLGPSPEGKELLPRAGFITLSIVMAMFTVPAIVLNSTVIVVSLMNKQLRQPLNY
ALVNMVAVADLGTALTGGVLSVVNNALGYFSLGRTGCIIEGFSVALFGITSLCTVALIAIERMFVVSPL
GPISFQTKHAVGGVALSWVWSLLWNTPLFGWGRYELEGVGTSCAPDWHNRDPNNVSYILCYFL
CFAVPFLIIVASYSKLILRQVSAMGCMEGGAAAKAEAKVARMVVLMMVVTFLISWLPYATLALVVVSH
PDAPISPLLGTVPVYLAKSSTVYNPLIYLYMNKQFRRYAVPFLLCGRDPWPSEEEGSEMQTTVATIN
NKVSPS

>Human Rhodopsin (NP_000530.1)

MNGTEGPNFYVPFSNATGVVRSPFEYPQYYLAEPWQFSMLAAYMFLIVLGFPINFLTYVTVQHK
KLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIER
YVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLAGWSRYIPEGLQCSCGIDYYTLKPEVNN

ESFVIYMFVVHFTIPMIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWVPYA
SVAFYIFTHQGSNFGPIFMTIPAFFAKSAAIYNPVIYIMMNKQFRNCMLTTICCGKNPLGDDEASAT
VSKTETSQVAPA

>*X. tropicalis* Parietopsin (NP_001039256.1)

MDGNSTTPGIAVNLTVMPTIFPRSGYSILSFLMFLNAVFSICNNAIVILVTLKHPQLRNPINIFILNLSFS
DLMMALCGTTIVVSTNYHGYFYL GKQFCIFQGFAVNYFGIVSLWSLTLLAYERYNVCEPIGALKLST
KRGYQGLVFIWLFCLFWAIAPLFGWSSYGPEGVQTSCSIGWEERSWSNYSYISYFLTCFIIPVGIIGFS
YGSILRSLHQLNRKIEQQGGKTNPREEKRVVIMVLFMVLAFLICWLPYTVFALIVVINPQLYISPLAAT
LPTYFAKTSPVYNPIIYIFLNKQFRITYAVQCLTCGHINLDSLEEDTESVSAQAENMLTPKTNQVAPA

>*Zebrafish* Parietopsin (BAL63415.1)

MENFAKTELTMMMVQPTIFPRMGYSILSYLMFINTTSLVFNNVLVIAVMVKNLHFLNAMTVIIFSLAV
SDLLIATCGSAIVTVTNYEGSFFLGDAFCVFQGFVAVNYFGLVSLCTLTLLAYERYNVCKPMAGFKLN
VGRSCQGLLLVWLYCLFWAVAPLLGWSSYGPEGVQTSCSLGWEERSWRNYSYLILYTL MCFVLPTA
IITYCYSNVLLTMRKINKSIECQGGKNCAEDNEHAVRMVLAMIIAFFICWLPYTAISVLVVVNPEISIPP
LIATMPMYFAKTSPVYNPIIYFLTNRKFRESSLEVLSCGRYISRETGGPLMGSSMQRGQSRVNPV

>*Iguana* Parietopsin (AB626970.1)

MENESSLVTEVAEGVTVRPTIFPRAGYGVLAFLMFLNALFSIFNNFLVIAVTLKNPQLRNPINIFILNLS
FSDLMMSLCGTTIVATNYHGYFYLGRRFCTFQGFVAVNYFGIVSLWSLTILAYERYNVVCQPLGTLQ
MSTKRGYQLLGFIVWFCLFWAVVPLFGWSSYGPEGVQTSCSIGWEERSWSNYSYLIVYFLSCFFIPV
LIIGFSYGNVIRSLHGLNKKVEQLGGKSNPEEEFRAVVMVLVMVVAFMICWLPYTVFALIVVFNPAL
NISPLAATIPTYLSKTSPVYNPIIYIFLNKQFRECAVEFITCGQVILTSQEEDISVSAIPDEGKALCKINQV
TPV

Appendix 6

Parapinopsin DNA Sequences

>Green_anole_pp

ATGGATTCTCTGGACACCAACACTTTGTACCTAATGCAAGCACAGTACGAGTGGTCTTGATGCTTGGTTAT
ACTATAATAGCAATTATTATGGCAACATCTTGCACTTTATCAGTAATTTTGAATACAGCCGTCATTGCCATAA
CTATTAAGTACAGGCAGCTTCGTCAGCCCATCAATTATTCTCTGGTTAATTTAGCAATAGCAGACCTTGGAGC
AGCTTTGTTAGGAGGAAGCCTGAATGTAGAAACAAATGCCGTTGGATATTACAACCTGGGCAGAGTTGGCT
GTGTCACCGAAGGTTTTGCTATGGCATTTTTTGGTATTGTGGCTTTGTGCACCATAGCAGTGATAGCTGTTGA
CCGCGCAATTGTAATCGCCAAGCCGATGGGAACAATTACATTTACTACCAGAAAAGCTATGATTGGTGTAGC
TGTGTCCTGGATATGGTCTCTTGTGTTGGAACACACCTCCTCTTTTTGGTTGGGGAGGATACCAGATGGAAGG
CGTTATGACCTCCTGTGCTCCAGATTGGGCTAACAGTGATCCTATCAATGTTTCCTACATTATCTGCTACTTTC
TCTTCTGCTTTACAATCCCATTCATCACAATCTTGGCCTCCTACGGTTACCTGATATGGACTCTGCGTCAGGTA
GCTAAAGTTGGACTAGCACAGCGAGGCTCAACAATAAGCGGAAGCACAAGTGTCACGGATGGTGATTG
TGATGGTTATGGCATTCTTGATCTGCTGGCTTCCTACGCAACTTTTGCCTTGGTAGTTGTGGGTAATCCTCA
AATCTACATTAATCCTATTATAGCGACCATCCCAATGTACATGGCAAAATCCAGCACATTCTATAATCCAATC
ATCTACATTTTTATGAACAAGCAGTTTCGCGATTGCCTTGTAAGGTGTCTCCTGTGTGGAAGAAATCCCTGTG
CTTCTGAGCAAACAGATGAAGATGATCTAGAAGTCTCTACCATTGCTCCTGCTCCTCCTCAAGACGTGGCA
AAGTTGCTCCTGTATAG

>Cat_Fish_pp

ATGGCATCCATTATTCTAATCAACTTTTCAGAGACTGACACACTGCATCTCGGCTCAGTAAATGATCACATCA
TGCCAAGAATAGGCTACACCATTTTGTCAATAATTATGGCTTTGTCTCCACTTTTGGCATCATCTTGAATATG
GTGGTGATTATTGTGACAGTGCGGTACAAACAACCTGCGTCAGCCACTAAACTATGCTCTGGTTAATCTAGCA
GTGGCTGATCTTGGCTGTCTGTGTTTGGGGGGCTACTCACAGCAGTCACCAATGCCATGGGCTACTTCAGT
CTTGGCAGGGTAGGATGTGTTTTGGAAGGATTTGCTGTTGCCTTCTTTGGTATTGCTGGTTTGTGCTCGGTCG
CAGTTATTGCTGTTGACCGCTACATGGTAGTATGCAGACCATTAGGTGCAGTGATGTTCCAAACCAAACATG
CCCTGGCAGGCGTGTTTTTCTGGGTCTGGTCCTTCATATGGAACACGCCACCACTGTTTGGATGGGGCA
GTTACCAGCTGGAGGGTGTGATGACCTCCTGTGCTCCCACTGGTACAGGAGAGACCCAGTCAATGTCTCCT
ATATCCTGTGCTACTTCATGCTCTGTTTTGCGCTGCCTTTGCCACCATTATCTTCTTACATGCACCTCCTTCA
TACCTTATGGCAGGTGGCTAAGTTGCAAGTGGCAGACAGTGGAAGCACGGCCAAGGTTGAGGTTCAAGTG
GCCCCGTATGGTGGTCATCATGGTCATGGCCTTTTTGCTTACATGGCTGCCATATGCAGCCTTTGCCCTGACTG
TGATCATTGATTCAAACATCTATATCAACCCAGTGATTGGCACAATACCCGCATACCTGGCCAAGAGCAGCA
CTGTCTTTAACCCCATCATTTATATTTTCATGAACAGACAGTTCAGGGACTATGCTTTGCCTTGTCTTTATGT
GGAAAAAACCTTGGGCAGCCAAAGAAGGACGAGACTCAGACACCAATACATTAACCTACTACAGTCAGCA
AGAACACCTCAGTGTCACCATTATAACCATAAGTATCACCGTTCATACCTAAAAAACGGGTGTTTCAAATATT
CAGCAATTCTCTTTCTATGTGTATTAATATTGTGAATATTTACAGTGTAATTTAGCTCTGGTATTTATTGTGC
TTGGATATTTATTTGGGTGAATGGTATTGTTTAAAGACATAGTCTTTAAGGTATTGTCTTGGATTATAGAAAT
ACAACATATTTTTTTTTAAAGCTTATCCCTGGCTACAATTTTCTCAAGAAATTGACAAATAAAGCATTGGTTA
TGCCCCCCCCAAAAAAAAGAAAGTTGAGCTCGAGTCCCAAGCCCGAATCTTGAGATTTCACCCACACCTGG
GGGGGC

>X.tropicalis_pp

TGTTGCCCTGACAATTTTCGGTGACTAAAAAGTCTCAGAACCAATGGCCGACGAGGCTCTTCTCCACCAAT
GATGAATGTGACCAATGAAGAGATGCACCCTGGGAAGGTTCTGATGCCCCGGATTGGTTACACCATTCTGG
CTTTAATAATGGCCGTATTCTGTGCCGCCGCTCTCTTTCTCAACGTCACGGTGATTGTGGTGACTTTCAAATAT
CGCCAACTGCGCCACCAATCACTACTCGCTGGTCAACCTGGCCATCGCCGACCTTGGCGTCACCGTGCTGG
GCGGGGCCTTGACAGTGGAGACAAATGCCGTGGGCTACTTTAACCTGGGGAGAGTGGGCTGCGTCATTGA
GGGTTTCGCCGTGGCATTITTTGGCATCGCGCTTTGTGCACCATTGCAGTAATAGCGTTGGACCGAGTGTT
CGTGGTGTGCAAGCCAATGGGCACACTCACATTCACCCCCAAACAAGCACTGGCCGGCATTGCTGCTTCCTG
GATCTGGTCGCTCATATGGAATACGCCCCGCTGTTTCGGCTGGGGTAGCTACGAGCTGGAAGGGGTGATGA
CATCCTGCGCCCCCACTGGTACAGCGCCGACCCAGTCAACATGTCCTACATCGTCTGCTACTTCTCCTTCTGC
TTTGCCATCCCATTCCTCATCATCGTGGGGTCTACGGGTACCTGATGTGGACTCTGCGGCAGGTTGCCAAGT
TGGGTGTAGCGGAAGGAGGTACAACGAGTAAAGCCGAGGTGCAAGTCTCCCGTATGGTGATTGTCATGAT
CTTGGCTTTCTGGTCTGTTGGCTCCCGTACGCCGCCTTTGCCATGACGGTTGTGGCAAATCCCGGAATGCAC
ATTGACCCCATATAGCCACCGTACCCATGTACCTGACCAAAACCAGCACCGTCTATAATCCAATTATCTACA
TTTTCATGAACAAGCAGTTCCAAGAATGTGTCAATCCCTTTCTGTTCTGCGGGAGGAATCCTTGGGCTGCTGA
AAAGTCAAGTTCTATGGAACTTCTATCAGCGTAACGAGTGGAACGCCTACGAAACGTGGCCAAGTGGCTC
CGGCGTAACGTGCTTTTATCTCCAGTATAACCGGC

>Rainbow_trout_pp

ATCATAACACAGGTGTCTGGCCAGGTGGGTAGTGCTCACCCTCAACTCCACTGCCCCAAAAGAAGACAGAA
TGGACCACCAACAGCTTCTCCCAATCTCCATGGAAACATCTCCTCCTCCCCGGGCTCGGTGAGTGAAGCGT
GCTCTCCAGGACGGGCTTACCATCCTGGCGGTGATCATTGGTGTGTTCTCGGTGTCCGGAGTCTGTATGAA
CGTCCTGGTCATCATGGTGACGATGAGACACCGAAAGCTGCGCCAGCCTTTGAACTACGCCCTAGTCAACCT
GGCAGTGGCTGACCTGGGCTGTGCTCTGTTTCGGGGGACTGCCACGATGGTCACCAACGCTATGGGCTACTT
CAGCATGGGGCGGCTAGGCTGCGTGCTGGAGGGTTTCGCTGTGGCTTTCTTTGGTATTGCAGGTCTGTGTTT
AGTGGCCGTAATAGCTGTGGACCGTTATGTGGTGGTGTGTCGACCAATGGGGGCGGTGATGTTCCAGACCA
GGCAGCGAGTCGGAGGGGTGGTCTGTCTGGGTGTGGTCGTTTTATGGAACACCCCCCTCTGTTTGGCT
GGGGCAGCTTTGAACTGGAGGGGGTTAGGACCTCCTGCTCCCCAACTGGTACAGCAGGGAACCAGGAAA
CATGTCTTACATCATCCTTTACTTCTGCTCTGCTTTGCCATTCTTTCTCCATCATCATGGTCTCCTACGCTCGC
ATCCTCTTCACTGCACCAGGTGTCCAAGCTGAAGGTGTTGGAGGGCAACAGCACCAACCCGCGTGGAGAT
ACAGGTTGTGCGTATGGTGGTAGTCATGGTGATGGCCTTCTGCTCAGCTGGTTGCCCTACGCGGCCTTCGC
CCTCTCTGTGATCCTTGACCCAGCCTCCACATCAACCCTCTCATTGCCACGGTTCCTCATGTACCTGGCCAAGA
GTAGCACTGTCTACAACCCATCATATACGTGTTTATGAACAGACAGTTCCGGGACTGTGCTGTTCTTTCTT
GCTGTGTGGTCTGAACCTTGGGCCTCAGAACCAAGTGGGCTCTGAAGCTGATACCGCCCTGTCTCCGTCAG
CAAGAACCCACGAGTCTCACCCCAATGACAGCAACTTTCTATTTTCTTGTACAATGTAAAGTTGCTGTTATA
AAATATAGTTTGATGTGCAATATTTGGTTGTAAGTGTTCGAAGAATACTAATAAAAGTATAGTCCATGATAT
ATGGACAGAGAC

>common_lguana

TTTCATAGGCATCTAAAGAAAAGATGACTGAAGACTGACAGACAGCATAAAACAACAAAGCTTGAGCAGT
ATAGGCATAAAAAAAGAGTATTTTATTTTGGTATCCTAACAACAAATGGATTTTTTGGACACCAGCACTTT
GTCAATGAATGCAACCACAGTACGAGTGGTCTTGATGCCCCGTATTGGTTATACTATAATAGCAATTATTAT

GGCAACATCTTGCACTTTATCAGTAATTTTGAATACAACCTGTCATTGCAGTAACTATAAAGTACAGACAGCTT
CGGCAGCCCATCAATTATCCCTGGTTAATTTAGCTATAGCAGACCTTGGAGCAGCTTTGTTAGGAGGAAGT
CTGAACGTAGAAACAAATGCTGTTGGGTATTACAACCTGGGCAGAGTTGGCTGTGTGACCGAAGGTTTTGC
TATGGCATTTTTTGGCATCGTGGCTTTGTGCACCATAGCAGTAATAGCTGTGACCGTGCAATTGTGATTGCC
AAGCCTATGGGAACACTAACTTTTACCACCAGGAAAGCAATGATTGGAGTCGCTATCTCCTGGATATGGTCT
CTTGTGGGAACACGCCTCCTCTTTTGGTTGGGGAGGATACCAGATGGAAGGCGTCATGACCTCCTGCGCT
CCAGATTGGTACAACAGTGATCCTATTAATGTCTCCTACATTGTCTGCTACTTTCTATTCTGCTTTACGATTCCC
TTCGTCACTATTTTGGTCTCCTATGGCTACCTGCTATGGACTCTGCGCCAGGTTGCTAAATTTGGAGTAACAC
AGCGAGGTTCAACAAACAAAGCGGAAGCTCAGGTGTCACGGATGGTGATAGTAATGGTTATGGCATTCTTG
ATCTGCTGGCTTCCTTATGCAACTTTTGCTTGGTAGTTGTGGGTAATCCTCAAATCTACATTAATCCTATTAT
AGCCACTGTCCCAATGTACATGGCAAAATCCAGCACATTTTATAATCCAATCATCTACATTTTATGAACAAG
CAGTTCCTGATTGCCTTGTGAGGTGCCTGCTGTGTGGAAGAAATCCCTGTGCTTCTGACCAAGCTGAGGAA
GATGATCTAGAAGTTTCTACCATTGCTCCTGCTCCTTCTCAAACGTGGCAAAGTTACTCCTATGTAGAAAA
CAATTCAGCATTGCTTTGGATTGATGTACATACAATGCTATGCACACTTTGTTCCCTGCCTGGAATCCCTGA
CACACTTACCCAGAAACAAGTCCCACTCTGTTCAAGTGAGGCTTAAATGTGTTAGCATTTCTGAATTTAATC
TGTCCTCAACTATTGTACAATGCCCTATCATTGTCTTTGAACCTGTGTTAATGTTAACTGTCCTTTACGGAACA
TTGTGTTATGGTGTATATAACATTCTGGAAATAGTTACACTGTTTGATACTCTT

>Zebrafish_pp_b

ATGGACTTTTTCCAGAATTCCTCAACGACTCTTCCACTCACTATGATGGCCCTGGTTTGAAGGTGCCTCTTCC
CCGGGCAGGATTCATCACGCTGTCCCTCCTCATGGCCGTGTTCTCCATCACATCTGTCGTGTTGAACGCTACA
GTGATAGTCGTCACCCTGCGGCACAAACAGCTGCGGCAGCCGCTCAATTTGCCCTGGTTAACCTCGCCGTG
GCCGACCTGGGCACCACACTAACAGGAAGCGTGCCATCTGTGGTGACCAATGCTGTAGGCTACTACATTATG
GGACGGATCGGCTGTGTATTGGAGGGATTTGTGTTGCATTTTTTGGTATCTCCGCGCTGTGCACCGTGGCT
CTGATCGCAGTGGAGCGTCTGTTTGTGGTGTGCAGACCTCTGGGCTCCATCACGTTTCAGTGCAGACACGCC
GCAGGAGGCCTGTTGTCTGCTGGCTTTGGTCTCTGATCTGGAACACTCCGCCTCTGCTGGGCTGGGGAAGC
TACCAGCTGGAGGGAGCGGGCACGTCCTGCGGGCCGCACTGGCAGAGTCGAGAACTCAGAGACGTGTCTT
ACATCATCTGCTACTTCTCCGTCTGCTTCGCGGTGCCATTCGCCATCATCCTGGTGTGCTATTATGCGCTGCTC
TACACGCTGAGACAAGTTTTACTGGAAATGGTAAACAATCTGGGAACTTTTGGCATGCTCTTTTACGCATAC
TACGATTTGGGACATATTTTCAAACACTATAAGCATTGCAACGCAGCTTTAACGTTAGTTTTATTTTTGTTGT
TGAATATCACTGGTTGCCTTACGCTGCTCTCGCTCTAACCGTCGCTCCTCAAACCTGAGGTCCAGCTGGCCGTG
CTGGTCAAGGTTTTACCCATATACATGGCCAAGAGCAGCACTGTATACAACCCCTTGATCTACATATACATG
AACAAACAGGTGAGCAGAACAAGCTGA

>Cherry_salmon_pp_b

GGGGGGAGTGTGGTGTGGAGGTTCTTACCATCTCCAAGAGAGTGAGAGCCTGTGTGGTGTGTGCTTGTGT
TGTTATATATCCATGATGCAGCTGCCAGCCTCTCTCCCAACGCTTCGTCCTACCTGGGACCATCCCCGGAGG
GCAAGGAGCTCCTGCCCCGAGCCGGCTTCATCACCTCTCCATCGTCATGGCTATGTTACCGTACCGGCCAT
TGTGTTGAACTCCACAGTGATCGTGGTGTGCTGATGAACAAGCAGCTGAGGCAACCTCTGAACTACGCCCT
GGTCAACATGGCCGTAGCTGACCTGGGGACAGCGCTGACTGGTGGGGTTCTGTCTGTGGTCAACAATGCTC
TGGGGTACTTCTCCCTGGGACGGACCGGCTGTATCATAGAGGGATTCTCTGTGGCTCTGTTTGGTATCACGT
CACTGTGCACGGTAGCTTTGATTGCCATAGAGAGGATGTTTGTGGTGAGCAAGCCGCTAGGACCAATCTCG
TTCCAGACCAAGCATGCGGTTGGGGGTGTGGCCTTGTGATGGGTTTGGTCTCTCCTCTGGAACACCCCGCCC

CTATTCGGCTGGGGAAGGTACGAGCTGGAGGGGGTGGGGACGTCGTGTGCCCCGGACTGGCACAACCGGG
ACCCCAATAATGTGTCATATATCCTGTGCTACTTCCTGCTGTGTTTTGCCGTGCCCTTCCTCATCATCGTGGCCT
CTTACTCTAAACTCATTTACACCTTAAGACAGGTCTCTGCAATGGGCTGTATGGAAGGGGGTGCAGCAGCCA
AGGCGGAGGCTAAGGTGGCCCGCATGGTGGTTCTGATGGTGGTTACTTTCCTGATCAGCTGGCTGCCCTAC
GCTACCCTGGCCCTTGTGGTGGTCTCCACCCCTGATGCCCCATTAGCCCTCTGCTGGGCACTGTGCCAGTCT
ACCTGGCCAAGAGCAGCACCGTGTACAACCCTCTCATCTACCTCTACATGAACAAACAGTTCCGGAGATATG
CAGTGCCCTTCCTCCTGTGTGGGAGGGACCCATGGCCATCTGAGGAGGAGGGGTGAGAGATGCAGACCACT
GTGGCAACCATCAACAACAAAGTCTCCCCAGTTGATCCATACAGGCATTTAGATAAATCCACTGATGGCAC
CAGCAGCACCAATAAAGTGTCACTCATATTCTATGCCTTTTTTAATACATTCATTAAATATACCTCTTAGTTTG
TAAGAATATGACTTATTAATACCTTAATAAGGAGATTGTATGACTTTCTTACCTGTGCACAACCAAAATAAC
TTTAGAGTATGGTTTTAATACTACATCCAGAATGGGATGAGAGACTGGCTGATCAGAGATATGTTACCTAC
AGTATGTTACAGAAGGTATATTAAGTGTGGAATCTATCGTTACTTTATGGTAAATATAAACACTATGAACAT
GCTGTTCAATACATAATGCATGGCTTATCTATAAACAAAGCACAAATAATGAACATACAGTGTAGAATTTCA
GTGTAAGTTTATTAGTAATAATCATTAGAAAAATAAAATATCATTATTGATTAATAATGAATTCAAGGCAAAA
AAAAAAAAAAAAAAAAAAAAA

>Lamprey_pp

GTGAGTGTATTTTTATAACTCGGTGCTTGATTTGAAGATCGTCTGTTCTGCAAGTAAGAGCCATGGAGAAT
TGACCTCGCTCGACCTCCTGCCAACGGCGAGGTCCCGTTGATGCCCCGCTACGGCTTCACCATCCTCGCCGT
GATCATGGCCGTGTTACCATCGCCTCGCTCGTCTCAACAGCACCGTCGTCATCGTCACCCTGCGCCACCGC
CAGCTACGCCACCCGCTCAACTTCTCGCTCGTCAACCTCGCCGTGGCGGACCTGGGCGTCACGGTGTTGCGC
GCCAGCCTCGTCGTGGAGACCAACGCCGTGCTACTTCAACCTCGGCCGCGTCGGCTGCGTCATCGAAGGGTT
CGCCGTGCTTTCTTCGGCATCGCCGCTCTGTGCAC
GATCGCCGTGATCGCCGTGATCGCTTCGTGGTGGTGTGCAAGCCGCTGGGCACGCTGATGTTACGCGGC
GCCACGCGCTGCTGGGCATCGCTGGGCTGGCTCTGGTCTGTTCTGTGGAACACGCCGCCGCTCTTCGGCT
GGGGCAGCTACGAGCTGGAGGGCGTGCGGACGTGCTGCGCGCCGACTGGTACAGCCGCGACCCCGCCAA
CGTGTCGTACATTACGAGCTACTTCGCTTCTGCTTCGCCATCCCCTTCCTCGTCATCGTGGTGGCGTACGGCCG
CCTCATGTGGACCTCCACCAGGTGGCCAAGCTGGGGATGGGCGAGAGCGGCAGCACCGCCAAGGCGGAG
GCGCAGGTGTGCGCATGGTGGTGGTCTGTTGGTGGCTTCTCGTCTGCTGGCTGCCCTACGCGCTCTTCG
CCATGATCGTGGTGACCAAGCCCGACGTGTACATCGACCCGGTCATCGCCACACTGCCATGTACCTGACCA
AGACGAGCACGGTCTACAACCCCATCATCTACATCTTCATGAACCGCCAGTTCGGGACTGCGCCGTGCCCTT
CCTGCTCTGCGGACGCAACCCCTGGGCAGAGCCTTCCTCGGAGTCGGCGACGGCGGCCTCCACGTGCGCCAC
CAGCGTGACCCTGGCGTCCGCGCCGGGCGAGGTCTCCCCGAGCTAGAGCGATCGATGAAT

>X. laevis pp

TGTTGCCCTGACAATTTTCGGTGACTAAAAAGTCTCAGAACCAAATGGCCGACGAGGCTCTTCTCCACCAAT
GATGAATGTGACCAATGAAGAGATGCACCCTGGGAAGGTTCTGATGCCCCGATTGGTTACACCATCTGG
CTTTAATAATGGCCGTATTCTGTGCCGCCGCTCTTTCTCAACGTCACGGTGATTGTGGTGACTTTCAAATAT
CGCCAACTGCGCCACCAATCACTACTCGCTGGTCAACCTGGCCATCGCCGACCTTGGCGTACCGTGCTGG
GCGGGGCCTTGACAGTGGAGACAAATGCCGTGGGCTACTTTAACCTGGGGAGAGTGGGCTGCGTCATTGA
GGGTTTTCGCCGTGGCATTTTTTGGCATCGCGCTTTGTGCACCATTGCAGTAATAGCGTTGGACCGAGTGTT
CGTGGTGTGCAAGCCAATGGGCACACTCACATTCACCCCAACAAGCACTGGCCGGCATTGCTGCTTCCTG
GATCTGGTCTGCTCATATGGAATACGCCCCGCTGTTTCGGCTGGGGTAGCTACGAGCTGGAAGGGGTGATGA
CATCCTGCGCCCCCAACTGGTACAGCGCCGACCCAGTCAACATGTCCTACATCGTCTGCTACTTCTCCTTCTGC
TTTGCCATCCCATCTCATCATCGTGGGGTCTACGGGTACCTGATGTGGACTCTGCGGCAGGTTGCCAAGT

TGGGTGTAGCGGAAGGAGGTACAACGAGTAAAGCCGAGGTGCAAGTCTCCCGTATGGTGATTGTCATGAT
CTTGGCTTTCTGGTCTGTTGGCTCCCGTACGCCGCCTTTGCCATGACGGTTGTGGCAAATCCCGGAATGCAC
ATTGACCCCATATAGCCACCGTACCATGTACCTGACCAAAACCAGCACCGTCTATAATCCAATTATCTACA
TTTTCATGAACAAGCAGTTCCAAGAATGTGTCATTCCCTTTCTGTTCTGCGGGAGGAATCCTTGGGCTGCTGA
AAAGTCAAGTTCTATGGAACTTCTATCAGCGTAACGAGTGGAACGCCTACGAAACGTGGCCAAGTGGCTC
CGGCGTAACGTGCTTTTATCTCCAGTATAACCGGC

>R. temporaria pp

TGTTGCCCTGACAATTTCCGGTGACTAAAAAGTCTCAGAACCAAATGGCCGACGAGGCTCTTCTCCACCAAT
GATGAATGTGACCAATGAAGAGATGCACCCTGGGAAGGTTCTGATGCCCCGGATTGGTTACACCATTCTGG
CTTTAATAATGGCCGATTCTGTGCCGCCGCTCTCTTTCTCAACGTCACGGTGATTGTGGTGACTTTCAAATAT
CGCCAACTGCGCCACCAATCAACTACTCGCTGGTCAACCTGGCCATCGCCGACCTTGGCGTCACCGTGCTGG
GCGGGGCCTTGACAGTGGAGACAAATGCCGTGGGCTACTTTAACCTGGGGAGAGTGGGCTGCGTCATTGA
GGGTTTCGCCGTGGCATTITTTGGCATCGCGCTTTGTGCACCATTGCAGTAATAGCGTTGGACCGAGTGTT
CGTGGTGTGCAAGCCAATGGGCACACTCACATTCACCCCCAAACAAGCACTGGCCGGCATTGCTGCTTCCTG
GATCTAGTCGCTCATATGGAATACGCCCCGCTGTTCCGGCTGGGGTAGCTACGAGCTGGAAGGGGTGATGA
CATCCTGCGCCCCCAACTGGTACAGCGCCGACCCAGTCAACATGTCCTACATCGTCTGCTACTTCTCCTTCTGC
TTTGCCATCCCATTCCTCATCATCGTGGGGTCCTACGGGTACCTGATGTGGACTCTGCGGCAGGTTGCCAAGT
TGGGTGTAGCGGAAGGAGGTACAACGAGTAAAGCCGAGGTGCAAGTCTCCCGTATGGTGATTGTCATGAT
CTTGGCTTTCTGGTCTGTTGGCTCCCGTACGCCGCCTTTGCCATGACGGTTGTGGCAAATCCCGGAATGCAC
ATTGACCCCATATAGCCACCGTACCATGTACCTGACCAAAACCAGCACCGTCTATAATCCAATTATCTACA
TTTTCATGAACAAGCAGTTCCAAGAATGTGTCATTCCCTTTCTGTTCTGCGGGAGGAATCCTTGGGCTGCTGA
AAAGTCAAGTTCTATGGAACTTCTATCAGCGTAACGAGTGGAACGCCTACGAAACGTGGCCAAGTGGCTC
CGGCGTAACGTGCTTTTATCTCCAGTATAACCGGC

> R. tigrina pp

TGTTGCCCTGACAATTTCCGGTGACTAAAAAGTCTCAGAACCAAATGGCCGACGAGGCTCTTCTCCACCAAT
GATGAATGTGACCAATGAAGAGATGCACCCTGGGAAGGTTCTGATGCCCCGGATTGGTTACACCATTCTGG
CTTTAATAATGGCCGATTCTGTGCCGCCGCTCTCTTTCTCAACGTCACGGTGATTGTGGTGACTTTCAAATAT
CGCCAACTGCGCCACCAATCAACTACTCGCTGGTCAACCTGGCCATCGCCGACCTTGGCGTCACCGTGCTGG
GCGGGGCCTTGACAGTGGAGACAAATGCCGTGGGCTACTTTAACCTGGGGAGAGTGGGCTGCGTCATTGA
GGGTTTCGCCGTGGCATTITTTGGCATCGCGCTTTGTGCACCATTGCAGTAATAGCGTTGGACCGAGTGTT
CGTGGTGTGCAAGCCAATGGGCACACTCACATTCACCCCCAAACAAGCACTGGCCGGCATTGCTGCTTCCTG
GATCTAGTCGCTCATATGGAATACGCCCCGCTGTTCCGGCTGGGGTAGCTACGAGCTGGAAGGGGTGATGA
CATCCTGCGCCCCCAACTGGTACAGCGCCGACCCAGTCAACATGTCCTACATCGTCTGCTACTTCTCCTTCTGC
TTTGCCATCCCATTCCTCATCATCGTGGGGTCCTACGGGTACCTGATGTGGACTCTGCGGCAGGTTGCCAAGT
TGGGTGTAGCGGAAGGAGGTACAACGAGTAAAGCCGAGGTGCAAGTCTCCCGTATGGTGATTGTCATGAT
CTTGGCTTTCTGGTCTGTTGGCTCCCGTACGCCGCCTTTGCCATGACGGTTGTGGCAAATCCCGGAATGCAC
ATTGACCCCATATAGCCACCGTACCATGTACCTGACCAAAACCAGCACCGTCTATAATCCAATTATCTACA
TTTTCATGAACAAGCAGTTCCAAGAATGTGTCATTCCCTTTCTGTTCTGCGGGAGGAATCCTTGGGCTGCTGA
AAAGTCAAGTTCTATGGAACTTCTATCAGCGTAACGAGTGGAACGCCTACGAAACGTGGCCAAGTGGCTC
CGGCGTAACGTGCTTTTATCTCCAGTATAACCGGC

Appendix 7

1 TGTGCCCCTGACAATTTCCGGTGACTAAAAAGTCTCAGAACCAAATGGCCGACGAGGCTCTTCTCCCACCA
71 ATGATGAATGTGACCAATGAAGAGATGCACCCTGGGAAGGTTCTGATGCCCCGGATTGGTTACACCATTCT
141 TGGCTTTAATAATGGCCGTATTCTGTGCCGCCGCTCTCTTTCTCAACGTACACGGTGATTGTGGTGACTTT
211 CAAATATCGCCAACTGCGCCACCCAATCAACTACTCGCTGGTCAACCTGGCCATCGCCGACCTTGGCGTC
281 ACCGTGCTGGGCGGGGCCCTTGACAGTGGAGACAAATGCCGTGGGCTACTTTAACCTGGGGAGAGTGGGCT
351 GCGTCATTGAGGGTTTCGCCGTGGCATTATTTTTGGCATCGCGGCTTTGTGCACCATTGCAGTAATAGCGTT
421 GGACCGAGTGTTTCGTGGTGTGCAAGCCAATGGGCACACTCACATTACCCCAAACAAGCACTGGCCGGC
491 ATTGCTGCTTCCTGGATCTGGTCGCTCATATGGAATACGCCCCGCTGTTTCGGCTGGGGTAGCTACGAGC
561 TGGAAGGGGTGATGACATCCTGCGCCCCAACTGGTACAGCGCCGACCCAGTCAACATGTCTACATCGT
631 CTGCTACTTCTCCTTCCTGCTTTGCCATCCCATTCTCATCATCGTGGGGTCTTACGGGTACCTGATGTGG
701 ACTCTGCGGCAGGTTGCCAAGTTGGGTGTAGCGGAAGGAGGTACAACGAGTAAAGCCGAGGTGCAAGTCT
771 CCCGTATGGTGATTGTCATGATCTTGGCTTTCTGGTCTGTTGGCTCCCGTACGCCGCCTTTGCCATGAC
841 GGTGTGGCAAATCCCGGAATGCACATTGACCCCATTTATAGCCACCGTACCCATGTACCTGACCAAAC
911 AGCACCGTCTATAATCCAATTATCTACATTTTCATGAACAAGCAGTTCCAAGAATGTGTCATTCCCTTTC
981 TGTTCTGCGGGAGGAATCCTTGGGCTGCTGAAAAGTCAAGTTCTATGGAACTTCTATCAGCGTAACGAG
1051 TGAACGCCTACGAAACGTGGCCAAGTGGCTCCGGCGTAACGTGCTTTTATCTCCAGTATAACCGGC

Primers Color Key

Fcdna

PNPd

PNPexon2 (O)

PNP1

PNPexon2 (I)